PROCEEDINGS
OF THE
NATIONAL ACADEMY OF SCIENCES
Volume 34 May 15, 1948 Number 5

GALACTIC AND EXTRAGALACTIC STUDIES, XIX. GIANT VARIABLE STARS OF THE LOOP NEBULA (30 DORADUS)

By Harlow Shapley and Virginia McKibben Nail

Harvard College Observatory

Communicated February 23, 1948

When in 1937 a large cluster of blue supergiant stars was found embedded in the heart of the 30 Doradus nebula by Shapley and Paraskevopoulos,[1] attention was called to the vast extent of the bright nebulosity that these stars excite, and also to the probability that heavy absorption prevails in the region, especially out beyond the luminous nebulosity. To test the degree and nature of this absorption, a study of the variable stars of the area was begun, and this communication reports on the results. An accidental by-product of the study—namely, the accurate determination for the first time of the absolute luminosity of giant Beta Lyrae variables—is of higher interest than the original goal.

1. The over-all diameter of the bright and dark nebulosity is something more than 30′. We have therefore chosen an area of one square degree, centered on the nebula, for special examination. In this area Miss Leavitt had already marked 38 stars as variable.[2] We now believe that five of them are not variable and five others are doubtful. Later Miss S. F. Mussells marked 26 additional objects in this square degree,[3] of which seven are certainly variable, ten doubtful and the others probably invariable.

Of the 36 stars[4] now definitely accepted as variable, only five lie within 15′ of the nucleus and inside the visible bounds of the nebulosity; but quite possibly dark absorption may extend irregularly over a much larger region and dim the stars throughout most of the chosen square degree. Of the 36 variables, 18 are Cepheids, 4 eclipsing binaries, 11 irregular, 1 long-period, 1 semiregular and 1 possibly cluster type.

A quantitative test for absorption is provided by the Cepheid variables. A total of 18 lie within the square degree. Their magnitudes have been based on the standards in Main Sequence VI and in two subsidiary sequences set up in the course of the extended photometric studies in the Large Cloud of forty selected Cepheid variables.[5] The magnitudes of

the 18 Cepheids in the neighborhood of the Loop Nebula are, therefore, on the same scale and zero point as those employed in determining the standard period-luminosity curve for the two Magellanic Clouds. Hence the standard curve can be used, through the agency of deviations, in testing for space absorption.

The numbers, positions, periods, maximum magnitudes and other data about all the variables of the 30 Doradus region are given in the accompanying table. Seven of the questionable stars are included, all with amplitudes of six-tenths of a magnitude or less. The periods are given in days, and the distances from the center of 30 Doradus (last column) are in minutes of arc.

The number and series of the photographic plates used for the measures have depended upon the position and brightness of each variable, and on the difficulties presented by its light curve. Some of the stars were troublesome because of overlapping star images; some because of the surrounding nebulosity. The differing background between variable and comparison star, and from plate to plate, makes the work in the 30 Doradus region comparable to work on the puzzling variables of the Orion Nebula, where many investigators during the past hundred years have been baffled by the difficulty of detecting regularity in the magnitude variations. In both the Orion and the 30 Doradus nebulosities many stars are suspected, and then dropped as unproved; the suspicion of variability is in part a consequence of the uneven nebulous background.

The proved variables around 30 Doradus, other than Cepheids, belong predominantly to the irregular class. The color of the variables has been determined only qualitatively from the intercomparison of red and blue plates made on the same night with the Bruce telescope with exposures of 120 minutes and 30 minutes, respectively. No dependable red magnitude standards are available for the Clouds. HV 2763 and HV 6002 were at minimum when the photographs were made, and in consequence below the plate limits.

The maximum magnitudes of all stars in table 1 range from 13.8 to 15.9, with a mean value of 15.06, which corresponds to absolute photographic magnitude -2.2 in the Cloud, with a spread from -3.4 to -1.3. The conspicuously red bright variables, like HV 2669, 2679, 2730, and 2798, if Cloud members, as is highly probable, must have absolute visual magnitudes in excess of -5, and therefore are even brighter than Betelgeux, which also is a red irregular variable.

The positions of the variables in table 1 are shown in figure 1, which is reproduced from a Bruce plate. The seven stars for which the type of variability could not be confirmed are not plotted. The following stars, all within the square degree, are now not considered appreciably variable on the basis of new measures of Bruce plates: HV 1015, 2693, 2753, 2755,

TABLE 1

HV	x′	y′	TYPE	PERIOD	MAX.	RANGE	MEDIAN	COLOR	R′
2669	14845	8846	Irreg.?	.....	14.3	0.5	...	Red	32
2674	14937	9572	Irreg.	.....	14.6	0.7	...	Red	26
5938	15042	11238	Cepheid	4.1743	15.5	0.8	15.9	..	33
2677	15066	10324	Irreg.	.....	15.0	1.5	...	Red	25
2679	15108	8926	Irreg.	.....	14.3	0.6	...	Red	28
5940	15148	9080	Cepheid	2.36993	15.9	0.7	16.25	..	26
2681	15156	9336	Irreg.	.....	14.6	0.6	...	Red	24
2685	15232	11364	Cepheid	6.5415	14.86	1.08	15.40	..	33
2687	15260	10930	Semi-reg.	73:	14.9	0.6	...	..	27
2691	15420	10674	Eclips.	1.2555	15.25	0.55	...	..	23
11088	15442	9994	Irreg.	.....	15.3	0.7	...	Red	18
5954	15461	8160	Cepheid	6.13595	15.51	0.56	15.79	..	33
2697	15467	11055	Cepheid	3.61885	15.2	1.1	15.75	..	26
1005	15689	10884	Cepheid	18.70976	13.96	1.46	14.69	..	22
1006	16035	8474	Cepheid	14.21141	14.25	1.50	15.00	..	24
2728	16113	8449	Irreg.	.....	14.5	1.1	...	Red	24
2730	16153	8714	Irreg.?	.....	14.2	0.6	...	Red	20
2732	16269	8484	Irreg.?	.....	14.5	0.6	...	Red	23
2740	16354	9601	Irreg.	.....	15.0	1.5	...	Red	5
2744	16553	8947	Eclips.	.....	15.2	0.5	...	..	15
2749†	16654	9021	Cepheid	23.0	15.2	0.9	15.65	..	14
5989*	16788	8130	Eclips.	1.3069	15.25	0.50	...	..	29
5987	16796	9910	?	.....	14.8	0.5	...	..	5
2761	16815	9024	Irreg.?	.....	15.3	0.5	...	Red	15
2763	16837	8066	L. Per.	400	14.5	>2.5	...	..	30
2765	16864	9261	Clust.?	.....	14.5	0.8	...	..	12
2773	17137	10945	Cepheid	6.34900	15.03	1.05	15.55	..	21
2774	17144	11038	Eclips.	3.65242	15.1	0.6	..	..	23
2776	17156	10864	Irreg.?	.....	15.3	0.5	...	Red	20
1017	17207	9064	Irreg.	.....	14.9	1.1	...	Red	17
2779	17225	11146	Cepheid	5.3217	15.25	0.75	15.62	..	25
2781	17254	9143	Irreg.	.....	14.7	0.9	...	Red	17
2787	17324	9946	Cepheid	11.442	14.48	1.03	14.99	..	15
6004	17339	10920	Cepheid	4.3704	15.9	0.7	16.25	..	24
2790*	17381	9588	Cepheid	6.60615	15.56	0.90	16.01	..	15
6002*	17484	9468	Irreg.	.....	15.9	>1.1	...	..	17
2793	17574	11304	Cepheid	19.1843	13.8	1.5	14.55	..	30
2795	17675	10752	Cepheid	3.91332	15.48	1.1	16.03	..	24
2798	17751	9629	Irreg.	.....	14.3	0.7	...	Red	21
6018	17880	8453	Cepheid	4.7833	15.7	0.6	16.0	..	33
2814	18245	9381	Cepheid	5.20879	15.2	0.9	15.65	..	30
6013*	18258	10668	?	.....	15.8	0.4	...	..	32
2815	18264	9033	Cepheid	11.98	14.4	0.9	14.85	..	33

* Positions recomputed.

† Period determined from 110 observations for years 1938 to 1946 is 23.1139 days.

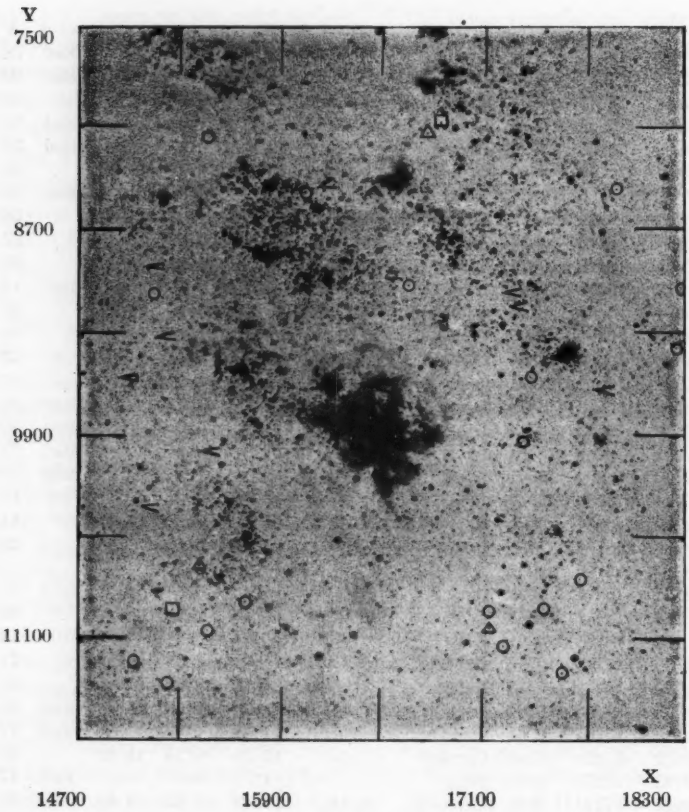

FIGURE 1

Negative of 30 Doradus. Variable stars are indicated as follows: open circles, Cepheids; triangles, eclipsing stars; squares, long-period and semi-regular; arrow heads, irregular; broken square, possibly cluster-type. Co-ordinate system in seconds of arc. The top 10′ strip on this photograph does not enter the investigation.

2770, 5930, 5931, 5935, 5950, 5952, 5955, 5960, 5968, 5970, 5993, 5994, 5999, 6005, 6007, 6014, 6021, 6024. Their considerable number again emphasizes the difficulty of variable star photometry in regions of irregular nebulosity.

2. Returning to the question of absorption, we plot in figure 2 the apparent median magnitude against the logarithm of the period for the 18 Cepheid variables. The straight line

$$\dot{m} = 17.14 - 2.08 \log P$$

where $\dot{m}$ is the apparent median magnitude and P the period in days, is the standard period-luminosity curve for the Large Cloud. The algebraic mean of the deviations in magnitude is O − C = +0.18; or, excluding HV 2749, it is +0.12. On the average, the absorption produced by the 30 Doradus nebulosity on these Cepheid variables is probably not more than one- or two-tenths of a magnitude. The outstanding indication of absorption is for HV 2749, for which O − C = +1.34. The star is nearest of all Cepheids to the nucleus of the nebula.[6]

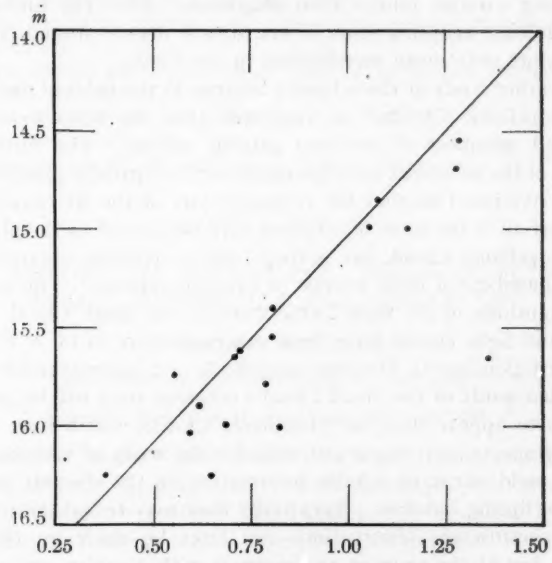

FIGURE 2

Deviations from standard period-luminosity curve indicate absorption by 30 Doradus nebulosity.

3. The four eclipsing binaries in the 30 Doradus square degree are remarkably similar:

HV	PERIOD, *d*	MAXIMUM, *m*	PRIMARY, *m*	SECONDARY, *m*
2691	1.2555	15.25	0.55	0.25
5989	1.3069	15.25	0.5	0.40
2774	3.65242	15.1	0.6	0.20
2744	.....	15.2	0.5	0.1:

The first three belong to the Beta Lyrae sub-class. For HV 2744 only four minima have been observed; the period may be related to 1.922 days. The chances are very good that all four eclipsing stars are members of the Magellanic Cloud and therefore are giant blue binaries with photographic absolute magnitude at maximum of approximately -2. This Cloud membership is made highly probable through a study of the frequency of faint variable stars in neighboring regions outside the Cloud—in VSF 212, 524 and 566.[7] One eclipsing binary for each ten square degrees is the find in these comparison regions where comparable magnitude intervals enter the systematic survey. In the 200 square degrees there is in fact no eclipsing variable fainter than magnitude 14.9. The finding of four 15th-magnitude eclipsing stars in one square degree around 30 Doradus can therefore only mean membership in the Cloud.

In an earlier study of the eclipsing binaries in the fields of the Large and Small Magellanic Clouds,[8] we suggested that the stars were probably superposed members of our own galactic system. The further study, however, of the frequency of eclipsing stars in comparable galactic latitudes has now convinced us that the eclipsing stars of the 30 Doradus region, and almost all of the score of eclipsing variables found on the plates of the Small Magellanic Cloud, are actually giant eclipsing binaries that are physical members of these nearest of external galaxies. The mean maximum magnitude of six Beta Lyrae stars in the Small Cloud, for which periods and light curves have been determined, is 15.13 ± 0.25 (m.e.), again corresponding to absolute magnitude -2, approximately.

A further study of the Small Cloud's eclipsing stars will be undertaken. It begins to appear that the Magellanic Clouds, which have heretofore in other respects been important tools for the study of various problems, can now yield our most reliable information on the absolute magnitudes of giant eclipsing binaries. Eventually they may reveal fainter eclipsing stars. Spectroscopic observations can later be made on these bright binaries; but at the moment we must report that spectra are unavailable for any of the 36 variables of the 30 Doradus region.

The very close similarity in magnitude and range of these bright binaries near 30 Doradus is reminiscent of the essentially identical magnitudes, ranges and periods of three long-period variables in the globular cluster 47 Tucanae.[9] In the Milky Way, eclipsing stars show large dispersions in periods, amplitudes and presumably in absolute magnitude at maximum.

4. In summary, we set out to test the degree of light absorption in the great 30 Doradus nebulosity of the Large Magellanic Cloud but ended with no very positive contribution to the solution of that problem. But three by-products are of considerable interest. The first is that 18 Cepheids in one square degree provide a good representation of the total Cepheid population in period-distribution, amplitudes, and relation of

luminosity to period. The second is that in this region of chaotic diffuse nebulosity, irregular variation in the light of deeply red stars is prevalent. And, of most significance, the absolute photographic luminosities of giant eclipsing binary stars are determined with relatively high accuracy, and a preference for absolute magnitude -2 is indicated.

[1] *Astrophys. J.*, **89**, 340–342 (1937); *Harvard Reprint* 141; see also *Harv. Circ.* 271, 8, 1925.

[2] *Harv. Ann.* **60**, No. 4, 87–108 (1908).

[3] *Ibid.*, **90**, No. 1, 1–25 (1933).

[4] A star only suspected of variability by Miss Leavitt now proves to be conspicuously variable and has been assigned the number HV 11088.

[5] *Harv. Ann.*, **90**, No. 10, 253–261 (1940); These PROCEEDINGS, **26**, 326–332 (1940); *Harvard Reprint* 202.

[6] These PROCEEDINGS, **26**, 541–548 (1940); *Harvard Reprint* 207; for HV 2749 see D. Hoffleit, *Harv. Bull.* 905, 25, 1937.

[7] *Harv. Ann.*, **90**, 171 (1934); *Harv. Bull.* 917, 1–5 (1943).

[8] *Harv. Bull.*, 916, 19–20 (1942).

[9] These PROCEEDINGS, **27**, 440–445 (1941); *Harvard Reprint* 228.

---

*A NEW THEORY OF SECONDARY NON-DISJUNCTION IN FEMALE DROSOPHILA MELANOGASTER**

BY KENNETH W. COOPER

DEPARTMENT OF BIOLOGY, PRINCETON UNIVERSITY

Communicated by A. H. Sturtevant, March 5, 1948

Sturtevant and Beadle[16] have shown that female exceptions constitute some 36.6% and 45.6%, respectively, of the female classes derived from *Drosophila melanogaster* mothers heterozygous for In(1)*C* or In(1)*dl-49* and possessing a *Y*-chromosome. The female "exceptions" are so-called because they derive both their *X*-chromosomes from their *XXY* mothers, not one *X*-chromosome from each parent as is usually the case. This anomalous situation is brought about by a failure of segregation, that is, by "secondary non-disjunction,"[1] of the *X*-chromosomes in a percentage of the oöcytic meioses, and is greatly enhanced by the presence of heterozygous *X*-chromosome inversions along with the *Y*-chromosome as Sturtevant and Beadle have demonstrated.

The frequency of non-disjunction is not equal to the frequency of exceptional females among the female class, however, for only half of the fertilized *XX* non-disjunctional eggs ordinarily emerge as adult flies, namely, the half that is fertilized by *Y*-bearing spermatozoa. The other half that is fertilized by *X*-bearing spermatozoa gives rise to superfemales having three *X*-chromosomes, and under ordinary conditions superfemales

only rarely eclose from their puparia.[9, 12] Accordingly, if p represents the *percentage* of secondary non-disjunction, or, better (XX)-(Y) segregation, and q the *frequency* of exceptional females, then

$$p = 100(2q/1 + q).$$

From this relation it follows that about 54% of all segregations in In(1)C/ $+/Y$ are non-disjunctional for the X-chromosomes, and some 63% of all segregations in In(1)*dl-49*/$+/Y$ females are also of the type (XX)-(Y). The high frequencies of secondary non-disjunction in these two cases are of more than passing significance, for as Sturtevant and Beadle[16] remark, they are in sharp disagreement with predictions based upon Bridges' generally accepted and now classic hypothesis concerning the mechanism of secondary non-disjunction.

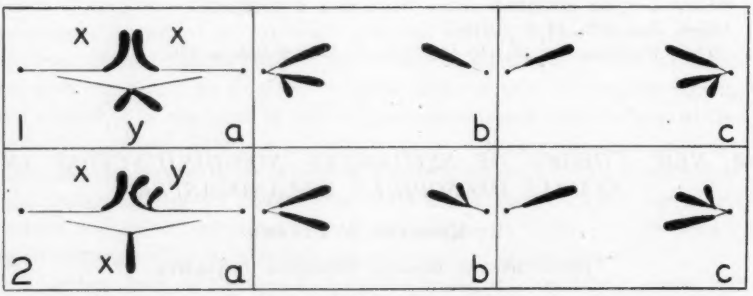

FIGURE 1

Purely diagrammatic representation of Bridges' (1916) hypothesis of secondary non-disjunction. Line 1, the consequences of XX-conjunction, Y being excluded from association. Line 2, the consequences of XY-conjunction, one X being excluded from association. Equally probable anaphase patterns arising from the particular metaphase associations are represented in blocks *b* and *c*. Secondary non-disjunction has a maximum value of 50% on this hypothesis.

Bridges' (1916)[8] hypothesis was in total accord with all of the known facts at the time at which it was proposed, and remained a satisfactory accounting of the mechanism and consequences of secondary non-disjunction until Sturtevant and Beadle's work upon heterozygous X-chromosome inversions. Bridges assumed that in XXY females the two X-chromosomes and the Y-chromosome are in mutual competition for a pairing or conjunctive partner, and that when any two of these three homologous chromosomes succeed in pairing, the third is left unassociated. Furthermore, the univalent chromosome was supposed to distribute randomly with respect to the poles, as is known to be the case for many univalents. The operation of the hypothesis is diagrammatically represented in figure 1. Line 1 represents the consequences of XX-pairing,

leaving Y a univalent to assort randomly. Line 2 represents the non-disjunctive process as Bridges conceived it. One X pairs with Y, the other X being a univalent. When anaphase occurs the univalent X goes to the Y-pole (2c) in 50% of the cases, and to the same pole of the spindle as the other X (2b) in the remaining 50% to give XX non-disjunctional nuclei. On Bridges' hypothesis, therefore, one hundred per cent pairing of Y with an X, could it be brought about, would give a maximum p of only 50%, and a maximum q of 0.33 as the frequency of exceptions among the female class of offspring. Thus the 54% non-disjunction found in In(1)$C/+/Y$ and 63% for In(1)*dl-49*$/+/Y$ cannot be accounted for on the basis of Bridges' assumptions. As Sturtevant and Beadle[16, 17] remark, a new interpretation of secondary non-disjunction must be sought as "the general problem of the mechanism of secondary non-disjunction is unsolved." It is the purpose of this note to provide such a new interpretation of the mechanism of secondary non-disjunction, along with data on several synthetic high non-disjunctional lines upon which the new interpretation is based.

All chromosomes (including the Y) of the *Drosophila melanogaster* stocks employed in the research now reported were derived from *Canton-S*, with the exception of the inverted X-chromosomes and the possible exception of the fourth chromosomes in some strains. Special methods were adopted to guarantee that homologous large autosomes were isosequential through all the stocks employed, and this was also the case for any non-inverted X-chromosome used in the crosses. The reason for building up such special stocks, the chromosomes of which are isosequential *inter se*, lies in Sturtevant's[14] notable discovery that heterozygous inversions in the large autosomes may markedly *decrease* rates of secondary non-disjunction in X-chromosome structural heterozygotes. Experiments on comparative rates of secondary non-disjunction may be expected to gain in sensitivity, to be more reproducible, and more revealing of unique qualities (if any) of inversions, therefore, when such potentially disturbing factors as autosomal structural (=sequential) heterozygosity are removed.[2]

The effects upon secondary non-disjunction of heterozygous combinations of In(1)B^{M1}, In(1)*dl-49*, and In(1)AM with wild-type (+) X-sequences, and in certain combinations with one another, are given for the female classes in table 1. (Data for the male classes are in good agreement with those for the female progeny, but are not recorded here as they require corrections for exchanges within inversions, as well as for viability differences in some cases.) In(1)B^{M1} is roughly ten map units long, and has its left break just to the right of *B* (57.0), and its right break probably to the right of *bb* (66.0). In(1)*dl-49* is perhaps thirty map units long, having its left break between *rb* and *cv* (i.e., between 7.5 and 13.7) and its right break between *fw* and *g* (between 38.3 and 44.4). In(1)AM is also about thirty

TABLE 1

THE PERCENTAGES OF EXCEPTIONAL FEMALES AND OF SECONDARY NON-DISJUNCTION IN *Drosophila melanogaster* MOTHERS HETEROZYGOUS FOR THE SPECIFIED INVERSIONS AND HAVING THE *Y*-CHROMOSOME AND ISOSEQUENTIAL AUTOSOMES DERIVED FROM *Canton-S*. THE STANDARD OR WILD TYPE SEQUENCE FOR THE *X*-CHROMOSOME IS REPRESENTED BY "+"

CONSTITUTION OF MOTHER	TOTAL FEMALE OFFSPRING	EXCEPTIONAL FEMALES, %	(XX)-(Y) SEGREGATIONS, %
+/+/*Y*(=control)	2121	0.9	2
In(1)B^{M1}/+/*Y*	2465	8	14
In(1)*dl-49*/+/*Y*	3705	54	70
In(1)*AM*/+/*Y*	2277	54	70
In(1)*dl-49*/In(1)B^{M1}/*Y*	1981	61	76
Ins(1)*dl-49*, B^{M1}/+/*Y*	5330	63	77
In(1)*dl-49*/In(1)*AM*/*Y*	2171	67	80

TABLE 2

THE OCCURRENCE OF EXCEPTIONAL FEMALES BY PRIMARY NON-DISJUNCTION IN *Drosophila melanogaster* MOTHERS HETEROZYGOUS FOR SPECIFIED INVERSIONS AND HAVING THE AUTOSOMES ISOSEQUENTIAL AND DERIVED FROM *Canton-S*. THE STANDARD OR WILD TYPE SEQUENCE FOR THE X-CHROMOSOME IS REPRESENTED BY "+." COMPARE WITH TABLE 1

CONSTITUTION OF MOTHER	TOTAL FEMALE OFFSPRING	TOTAL EXCEPTIONAL FEMALES	(XX)-(O) SEGREGATIONS, %
+/+ (=control)	8081	1	0.02
In(1)B^{M1}/+	2660	0	..
In(1)*dl-49*/+	3155	10	0.6
In(1)*AM*/+	1133	1	0.2
In(1)*dl-49*/In(1)B^{M1}	1389	4	0.6
Ins(1)*dl-49*, B^{M1}/+	5613	10	0.4
In(1)*dl-49*/In(1)*AM*	1159	0	..

map units in length, extending from near *lz* (27.7) to between *B* and *Bx* (i.e., between 57.0 and 59.4).[3] It is seen that the inversion heterozygotes markedly differ in non-disjunctional rates from the control, and that secondary non-disjunction sharply rises with an increase in the structural difference existent between the two *X*-chromosomes. Furthermore, excepting the cases of +/+/*Y* and B^{M1}/+/*Y*, none of the non-disjunctional rates found are interpretable upon the basis of the classical hypothesis which sets 50% as the uppermost limit of secondary non-disjunction.

Table 2 records the occurrence of primary exceptions among the offspring of females having *X*-chromosome constitutions identical with those of table 1, but lacking a *Y*-chromosome. As Sturtevant and Beadle[16] concluded from their own data, and as table 2 shows, mere heterozygosity for *X*-chromosome inversions has but a negligible effect upon the final outcome and regularity of *X*-chromosome segregation in female *Drosophila melanogaster*. Thus very high non-disjunction rates depend not only upon

the particular combination of heterozygous inversions present, but in addition *require* the presence of a *Y*-chromosome. There is no doubt that Bridges was correct in his opinion that the *Y*-chromosome somehow brings about non-disjunction by its own segregative involvements.

Now it is known that the presence of these inversions in heterozygous combinations markedly increases the proportion of non-exchange tetrads formed by the *X*-chromosomes,[10, 13, 15, 16] whether or not a *Y*-chromosome is present. Indeed, as a conservative estimate, at least 76% of the *X*-chromosome pairs in Ins(1)*dl-49*, $B^{M1}/+$, *Dp*(1; 1)*112* are non-exchange tetrads,[10] and this is also the case where the uninverted chromosome lacks a duplication. The degree to which crossing-over is eliminated may be taken as one measure of the meiotic dissimilarity of the *X*-chromosomes caused by their sequential non-homology in principally euchromatic lengths.

But as the *X*-chromosomes become sequentially dissimilar through the introduction of heterozygous combinations of inversions affecting their euchromatic lengths, so they become effectively more like *Y*-chromosomes in their pairing affinities. A moment's reflection will bring conviction that were the euchromatic lengths of two *X*-chromosomes to be made wholly dissimilar, but the so-called "inert" or chromocentral regions to remain essentially unaltered, then conjugation between the *X*-chromosomes would be predominantly a heterochromatic affair—a process occurring almost exclusively between the chromocentral regions.[4] This, however, is merely another way of stating that conjunction becomes more frequent in those regions of *X* wherein pairing with *Y* normally occurs. Hence, as is found to be the case, when a *Y*-chromosome is present non-disjunction markedly rises in *X*-chromosome inversion heterozygotes, probably because the sequential differences between the *X*-chromosomes in effect emphasize their likenesses to the *Y*-chromosome. Thus, where the *X*-chromosomes are everywhere in sequential homology, *Y* induces a rise in non-disjunction from 0.02% in $+/+$ to a mere 1.6% in $+/+/Y$. In other words the *X*-chromosomes share more similarities with each other than they do with the *Y*-chromosome, and *Y* becomes an effective partner to *X* in only a small proportion of the cases—perhaps less than 2% of all associations in $+/+/Y$. But where there is a great reduction in sequential homology between the two *X*-chromosomes, as in the combination In(1)*dl-49*/In(1)*AM*, non-disjunction rises from an indeterminately small fraction in *dl-49/AM* to about 80% when merely a *Y*-chromosome is added to the karyotype. *Y*, in other words, is a more attractive partner to either *X*-chromosome in this combination than either *X*-chromosome is to the other. The apparent paradox that each *X* in effect finds *Y* a more attractive and frequent conjunctive partner than the other *X*-chromosome, although the *X*-chromosomes remain alike in the very regions with which

they share homology with Y, is resolved by the assumption that *both arms* of the Y-chromosome share homology and may thus conjoin with an X.[5] Hence if one X pairs with one arm of Y, then the other X must conjoin with the other arm of Y if it is to have a meiotic partner at all. The mechanical aspects of this mode of bringing about an apparent non-disjunction by trivalent formation of both X-chromosomes with a Y-chromosome are diagrammed in figure 2. Since, for example, X_a pairs with arm Y_a, and X_b with Y_b, the segregation patterns are $X_a - Y_a$ and $X_b - Y_b$. But in an alternately oriented trivalent Y_a and Y_b perforce must go to one pole, since they possess the same kinetochore, and X_a and X_b undergo forced segregation to the other pole since each disjoins from Y.

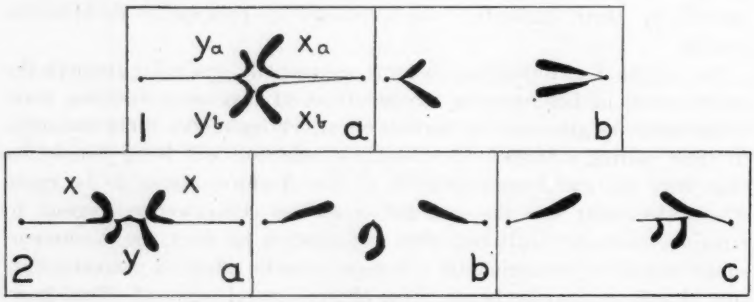

FIGURE 2

Purely diagrammatic representation of the interpretation of secondary non-disjunction proposed here. *Line 1 of figure 1* is held to represent accurately the consequences of XX-conjunction in XXY females. Line 1 of figure 2 shows an XYX trivalent and the imposed non-random segregation following alternate coörientation of the associated chromosomes. Line 2 represents a linear orientation of an XYX trivalent, and two of the possible consequences in anaphase. Such misoriented trivalents may lead to the loss of Y and other irregularities, but do not give rise to exceptional gametes. The majority of trivalents are supposed to orient alternately as in 1*a*, hence this interpretation admits a possible maximum of 100% secondary non-disjunction.

The interpretation of secondary non-disjunction that is offered here asserts, therefore, that secondary non-disjunction is not a case of non-disjunction at all, but an example of imposed non-random segregation from a sex-chromosome trivalent.[7] The mechanism proposed has a direct analogy with the segregative mechanism normally occurring for X_1, X_2 and Y in the spermatogenesis of *Drosophila miranda*.[11] It may be pointed out that XYX trivalents formed at meiosis in any XXY female of *Drosophila melanogaster* may be greater than the percentage of secondary non-disjunction owing to the probable occurrence of some linear orientations of the trivalent at first metaphase (Fig. 2, line 2).

The principal virtue of this new interpretation of secondary non-disjunction lies not alone in the fact that it wholly accounts for data not

resolvable by Bridges' hypothesis, but that it can, if false, be disproved by a direct test of its predictions. The scheme proposed predicts that (1) secondary non-disjunction must rise on the average as the euchromatic sequences of the two X-chromosomes become increasingly different; (2) the presence of two Y-chromosomes should decrease the rate of secondary non-disjunction; (3) either the long arm or short arm of Y alone could not regularly give greater than 50% secondary non-disjunction; and (4) that in those cases where secondary non-disjunction exceeds 50%, the X-chromosomes must have a partner that can pair with both X-chromosomes simultaneously (i.e., in general the partner must have two arms). And so forth.

Each of these implications has been put to test, and each has so far proved to be in accord with the experimental facts. For example, (1) tests of $+/+/Y$, In(1)$sc^7/+/Y$, In(1)$B^{M1}/+/Y$, In(1)sc^7/In(1)B^{M1}/Y and Ins(1)sc^7, B^{M1}/In(1)*dl-49*/Y give as secondary non-disjunction rates: 2%, 14%, 14%, 63% and 78%, respectively. This is the very order that would be predicted, and since In(1)sc^7 and In(1)B^{M1} are of the same order of size, close similarity in their effects on raising the secondary non-disjunction rate is to be expected.

(2) On the new interpretation, if each arm of Y is equivalent to X in its pairing, and if the order of pairing is random, but pairing complete in so far as possible, then in XXY females secondary non-disjunction would be expected to approximate a p of 80%. It is interesting to note that this is the uppermost value for secondary non-disjunction so far attained in my experiments. In $XXYY$ females tetravalents may be expected if corresponding arms of Y may pair, but their characteristics of orientation is an unknown factor. However, it may be assumed that, for necessity of spatial economy on the spindle, alternate orientations (N) should exceed adjacent (M). If results for equality of these two classes of orientation be calculated, as well as for total alternate orientation, then the expectation is that secondary non-disjunction in $XXYY$ females will have a p with some value in the interval from 27 to 52%. If it be assumed that tetravalents orient wholly at random, then p should approximate 52%. Should it be the case that tetravalents never form, then p should approximate 42%. The three possibilities treated above cover the more probable conjectures about conjunction in $XXYY$ females, and the calculated values are closely similar to or include the experimental value of 36% obtained from Ins(1)*dl-49*, $B^{M1}/+/Y/Y$.[6] In the control, Ins(1)*dl-49*, $B^{M1}/+/Y$, p is 78%.

(3) An effectively one-armed Y^s-chromosome obtained through the kindness of Professor H. J. Muller gives about 24% non-disjunction in Ins(1)*dl-49*, $B^{M1}/+/Y^s$ whereas the control crosses of Ins(1)*dl-49*, $B^{M1}/+/Y$ give 77%. (The Y^s-chromosome has not yet been gotten into

a background of *Canton-S* autosomes, so this test with Y^s may have given a somewhat low value. Internal evidence supplied by the controls, which are XXY daughters of the XXY^s females, suggests, however, that the true effect of the Y^s-chromosome on secondary non-disjunction in this X-chromosome karyotype will not be found to differ appreciably from the low value so far obtained.)

(4) All Y-chromosomes in the experiments undertaken which give non-disjunctional rates over 50% (with heterozygous Ins(1)*dl-49*, B^{M1}) have two effective arms. These Y-chromosomes include Y^{cs}, Y^{sw-b}, $sc \cdot Y^L$, and others.

The interpretation here posed, therefore, is in agreement with a wide variety of new facts, as well as with those past discoveries which make up the body of available data on secondary non-disjunction in *Drosophila melanogaster* females.

Summary.—A new theory of secondary non-disjunction in female *Drosophila melanogaster* is proposed. It asserts that the process is actually a case of imposed non-random segregation from an XYX trivalent. Each arm of Y is assumed to pair with a different X-chromosome to form a trivalent which usually orients in alternate pattern at first meiotic metaphase. Thus the cytological features are envisioned as similar to the directed segregation from X_1YX_2 trivalents in the spermatogenesis of *Drosophila miranda*. Previously known facts, as well as results of genetic experiments designed to test the scheme, appear to be in full accord with this interpretation.

* The research reported here was first undertaken in 1945 during tenure of a John Simon Guggenheim Memorial Foundation Fellowship at the Kerckhoff Biological Laboratories of the California Institute of Technology. The experiments have continued uninterruptedly since then in the Biological Laboratories of Princeton University and, during the summers of 1946 and 1947, at the Marine Biological Laboratory, Woods Hole, Massachusetts.

[1] Failure of the two X-chromosomes to segregate to opposite poles in the oögenesis of an XX-female is referred to as "*primary* non-disjunction." Inclusion of both X-chromosomes in a single polar group at meiosis of an XXY-female is referred to as "*secondary* non-disjunction." The distinction is an important one, for the mechanisms of primary and secondary non-disjunction are wholly dissimilar, the latter being dependent upon the extra Y-chromosome.

[2] Sturtevant and Beadle[16] did not deliberately clear their stocks of possible sequential disturbances of the autosomes, for no such effect of autosomal inversions upon secondary non-disjunction had been anticipated in 1936. This may well account for the significantly lower rate (63%) found by them than that (70%) reported here for secondary non-disjunction in In(1)*dl-49*/ + /*Y*. *Cf.* table 1.

[3] Additional information on these inversions can be had from: Bridges, C. B., Carn. Inst. Wash. Pub. No. 552, viii + 257 pp. (1944).

[4] This conclusion perhaps deserves a general discussion of the nature of synapsis. Suffice it to say in this connection that the idea of specific forces between homologous genes is without proof, and is almost certainly not a necessary hypothesis.

[5] Proof that each arm of *Y* shares homology with *X* will be found in Neuhaus's work; *v.* Neuhaus, M., *Genetics*, **22**, 333–339 (1937).

[6] A noteworthy outcome of these experiments, that will be reported in detail in another paper, is that, in general, flies whose eyes are $+/+$, $+/w^a$, w^a/w^a, $+/v$, v/v, v/w^a, *v/car*, *car/car*, etc., and which possess two (or more) *Y*-chromosomes beyond normal for the karyotype (i.e., *XXYY*♀♀ and *XYYY*♂♂) have flecked and/or mottled eyes.

[7] Bauer suggests that the union of two *Y*-chromosomes to form a V made possible the general elimination of $(X_1Y_1) - (X_2Y_2)$ segregations in tetraploid earwigs. A mechanism similar to the one proposed here would result by such translocation, and give principally $(X_1X_2) - (Y_1Y_2)$ segregations at spermatogenesis. *v.* Bauer, H., *Zeitschr. Naturforsch.*, 2b, 63–66 (1947).

[8] Bridges, C. B., *Genetics*, **1**, 1–52, 107–163 (1916).

[9] Bridges, C. B., *Science*, **54**, 252–254 (1921); *Am. Naturalist*, **56**, 51–63 (1922); *ibid.*, **59**, 127–137 (1925).

[10] Cooper, K. W., *Genetics*, **30**, 472–484 (1945).

[11] Cooper, K. W., *Ibid.*, **31**, 181–194 (1946).

[12] Dobzhansky, Th., these PROCEEDINGS, **14**, 671–675 (1928).

[13] Gershenson, S., *Jour. Genetics*, **30**, 115–125 (1935).

[14] Morgan, T. H., and Sturtevant, A. H., Carnegie Inst. Wash. Year Book, **43**, 164–165 (1944).

[15] Stone, W., and Thomas, I., *Genetica*, **17**, 170–184 (1935).

[16] Sturtevant, A. H., and Beadle, G. W., *Genetics*, **21**, 554–604 (1936).

[17] Sturtevant, A. H., and Beadle, G. W., *An Introduction to Genetics*, Saunders Co., Philadelphia, 1939, pp. 1–391.

---

*THE ORIGIN OF VOLUTIN ON THE CHROMOSOMES, ITS TRANSFER TO THE NUCLEOLUS, AND SUGGESTIONS CONCERNING THE SIGNIFICANCE OF THIS PHENOMENON**

BY CARL C. LINDEGREN

SOUTHERN ILLINOIS UNIVERSITY, CARBONDALE, ILLINOIS

Communicated by Leo Loeb, March 27, 1948

Volutin as a Cytological Entity.—Volutin, which is widely distributed in bacteria and fungi and is also called metachromatin, is a clearly defined cytological entity, distinguished by its ability to retain basic stains when the cells are rinsed in dilute acid. When volutin is stained with methylene blue or toluidine blue, it retains the stain after treatment with 1 per cent sulphuric acid. In addition to being "acid fast," volutin has the characteristic of staining red or purple with methylene blue or toluidine blue, in contrast to other basophilic components of the cell which stain blue with these dyes. Both metaphosphates and estersulphates retain dyes after destaining and stain red or purple with toluidine blue. Volutin is widely distributed in fungi and bacteria, while the estersulphates[1] are apparently widely distributed in higher animals. Wiame's work[2] strongly suggests

that the volutin in yeast cells is metaphosphate, because the addition of the phosphate to the cell increases the amount of volutin demonstrable in the cell, and extracts from these heavily stained volutin-containing cells yield reasonably large quantities of metaphosphate. There is, however, no cytological means of distinguishing between metaphosphates and estersulphates.

The present paper deals with the cytological observations on the cycle through which volutin originates on the chromosomes and is transferred to the nucleolus. A hypothesis concerning the significance of volutin is also presented. This hypothesis probably depends for its final validity upon the demonstration of metaphosphate in the animal cells, but this has not yet been achieved.

The Nucleolus of the Yeast Cell.—Lindegren[3] was the first to describe the chromosomes in the vacuole of the yeast cell. Lindegren and Lindegren[4] described the mitosis by which these chromosomes are partitioned equally to mother and daughter yeast cells. Rafalko[5] showed that the chromosomes in the nuclear vacuole of the yeast cell are Feulgen-positive. Prior to his work, it was argued that the chromosomes were the Feulgen-positive structures in the body attached to the nuclear vacuole which Lindegren and Lindegren described as the centrosome.

Rafalko (unpublished) has identified the nucleolus as a structure, staining with light green counterstain attached to Feulgen-positive chromosomes in the nuclear vacuole of the yeast cell. We had previously described the nucleolus, but had called it "balled-up" chromosomes. Rafalko's demonstration of it as a Feulgen-negative globular structure attached to the Feulgen-positive chromosomes has disclosed its true nature. We had previously designated the cytoplasmic granules as the nucleolar equivalents, following Caspersson and Brandt's theory[6] of their significance. However, the discovery of the true nucleolus makes this invalid. Except for the excessive size of the centrosome and its high content of desoxyribosenucleoprotein, the yeast cell is quite conventional. The cytoplasmic granules are the conventional mitochondria.

The Transfer of Volutin from Chromosomes to Nucleolus.—Chromosomes and nucleolus can be demonstrated with a stain specific for volutin[7] which contains formaldehyde, as a killer and fixative, and locates the volutin in the cell. If the chromosomes are coated with volutin, they are stained purple. If the nucleolus is coated with volutin, it is likewise stained purple. As soon as either the chromosomes or the nucleolus lose their volutin, they are no longer specifically stainable. The density of the stain also gives some idea of the quantity of volutin present. The best results are obtained when the stain is adjusted to a pH of 2.5, which prevents the proteins in the cytoplasm from also becoming stained. The stain contains 4 cc. saturated aqueous toluidine blue, 20 cc. formalin and sufficient acetic

acid to bring the pH to 2.5. A loopful of cells is placed in 10 drops of the stain in a 4-inch test tube. The tube is allowed to stand overnight, centrifuged, the supernatant fluid discarded and a loopful of the sediment is mounted on a slide, covered with a cover glass, pressed down with blotting paper and sealed with paraffin.

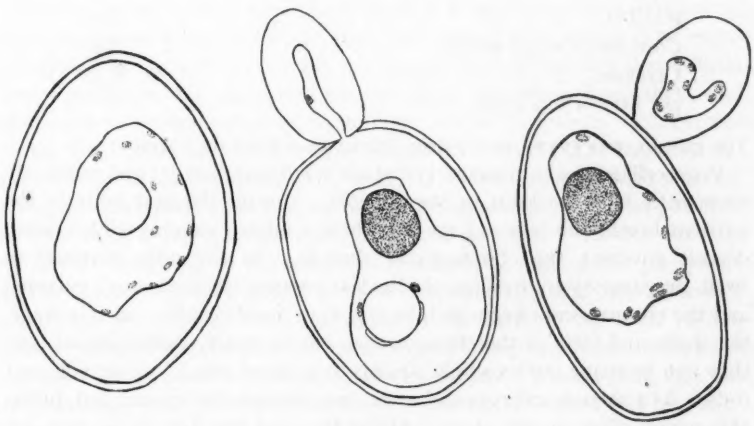

FIGURE 1

Yeast cells stained for volutin; the vacuole is shrunken by fixation. (*a*) A single large unstained nucleolus with many small, light pink peripheral chromosomes. (*b*) A budding cell with one stained and one unstained nucleolus. Two stained and two faintly stained chromosomes are attached to the unstained nucleolus. (*c*) A budding cell with a deeply stained nucleolus and very faintly stained peripheral chromosomes. No nucleolus visible in bud.

The growth of yeast was followed closely from resting to actively growing, vigorously fermenting cells. In resting cells, the chromosomes are unstained and the nucleolus is practically invisible. Before budding begins, the chromosomes become coated with volutin, and two unstained nucleoli become visible in each cell. As budding begins, the nucleoli increase in size, volutin disappears from the chromosomes and appears on the nucleolus. Transfer of volutin occurs by a direct attachment of *chromosome to nucleolus*, for two or three small, dark purple dots can be seen attached to an unstained nucleolus. Generally, one nucleolus becomes coated with volutin before the other, and subsequently both fuse. These fused nucleoli are the commonly described refractile "dancing bodies" in the yeast vacuole. The nucleolus is a large, round structure, possibly 2 or 3 microns in diameter. At first the interior of the nucleolus is not deeply stained with the purple dye; the color is on the outer edge or surface of

the structure. The nucleolus often becomes quite large before volutin appears on it.

The following typical results were obtained when 7 g. of compressed yeast were inoculated into 500 cc. of medium of the following composition:

Liquid yeast extract	6 cc.
Peptone	5 g.
KH_2PO_4	2 g.
Corn steep water solids	2 g.
Cerelose	40 g.

per liter of tap water

The culture was grown in a 6-liter Erlenmeyer flask on a shaker.

Yeast cells in a cake of baker's yeast are small and compact and ordinarily show little or no volutin in the vacuole. During the first hour in the nutrient broth, the cells enlarge and show a faintly staining pink volutin deposit covering the chromosomal threads. As the cells continue to swell, preparatory to division, the nuclear vacuole becomes much enlarged and the chromosomes begin to take the stain more deeply. In this stage, the shape and form of the chromosomes can be clearly distinguished, and they can be made out as single strands or pairs of small, rather compact rods. As a vacuole enlarges still more, two nucleoli also appear, but during this period, they do not stain. About the time the first buds come up, one or two nucleoli are visible as large, light green, refractile bodies in the nuclear vacuole. The chromosomes now appear as a large number of individual bodies on the inner periphery of the vacuole. After a while, one nucleolus takes on the volutin stain, followed shortly by the other, or both fuse. At the same time, the color disappears from the chromosomes or becomes very much diminished. When the bud is growing at its most rapid rate, the nucleolus is swollen and densely stained, showing that it contains a large quantity of volutin. The chromosomes are visible as small, faintly stained peripheral bodies on the inner wall of the nuclear vacuole. The nucleolus is usually visible as a stained structure only in the vacuole of the mother cell, but in some cases the vacuole of the daughter cell also contains a stained nucleolus. After division has ceased, the nucleolus retains a dense coating of volutin, but the chromosomes are unstained.

"Abnormal" Distribution of Volutin in the Yeast Cell.—Wiame[2] has shown that yeast cells can be starved for phosphate by growing them in a medium deficient in phosphate, such as beet molasses. When growth finally ceases, chemical analysis reveals no metaphosphate in the cells, and the cells do not contain volutin. Furthermore, the cytoplasm is colorless showing a marked diminution of basophilic protein in the cytoplasm. He centrifuges these cells, washes them in water and re-suspends them in a 2 per cent sugar solution, adding $M/30$ KH_2PO_4. Wiame kindly performed this experi-

ment with me in my laboratory and allowed me to follow the process cytologically. The chromosomes became rapidly coated with volutin, staining red by the volutin stain. Volutin sometimes appeared on the chromosomes within three minutes after the addition of the phosphate. After twenty minutes, volutin began to appear in the cytoplasm and the volutin in the vacuole decreased rapidly as the entire cytoplasm became faintly pink. Wiame's chemical tests show that there is a steady increase in the total metaphosphate in the cells and that the over-all concentration of metaphosphate in the cells continually increases in spite of the fact that volutin first appears in the chromosomes and then disappears from them and continues to accumulate later exclusively in the cytoplasm. Concurrent stains with the methylene blue dead cell stain reveal that the cells containing volutin in the cytoplasm are dead. This is consistent with the fact that metaphosphate is capable of precipitating proteins. The appearance of volutin in detectable amounts in the cytoplasm is a definite criterion of death. The *rate* of death can be reduced by adding the phosphate to the stained cells very slowly, but in a medium containing only 2 per cent sugar, the phosphate eventually results in the death of the cells. When normal cells (in an adequate medium) take up phosphate, the volutin stain reveals that as growth progresses, the cytoplasm becomes blue and the chromosomes red, indicating that there is basophilic protein but no stainable volutin in the cytoplasm of the healthy yeast cell. Healthy growth occurs only if a properly balanced nutrient is added slowly; phosphate alone produces a condition of imbalance which eventually results in death.

Volutin and Ribosenucleoprotein Synthesis.—The loss of stain from the chromosomes and its accumulation in the nucleolus indicate that the volutin formed on the chromosomes has been transferred to the nucleolus. Apparently, the synthesis of volutin occurs on the surface of the chromosome, and as soon as the chromosome becomes densely loaded, the transfer to the nucleolus occurs. Caspersson and Schultz[5] have described the nucleolus of higher plants and animals as the center for the synthesis of ribosenucleoprotein. They showed that after ribosenucleoprotein is synthesized in the nucleolus in the cell of a higher plant or animal, it travels through the membrane into the cytoplasm. Caspersson and Brandt[6] did not observe the nucleolus in the yeast cell with ultra-violet light, possibly because the cytoplasm contains so much nucleoprotein that the smaller amount in the nucleolus is overshadowed. They did show (with ultra-violet light) that ribosenucleoprotein accumulates on the outside of the nuclear vacuole. They thought it was synthesized at this point, but the demonstration of a conventional nucleolus in the nuclear vacuole suggests that the accumulation of ribosenucleoprotein, demonstrated on the surface of the nuclear membrane, may be the accumulation of this substance, as it diffuses from the nucleolus through the nuclear membrane

and into the cytoplasm. This view is in agreement with Caspersson and Schultz's ideas[8] concerning the rôle of the nucleolus in the synthesis of ribosenucleoprotein in higher plants and animals, but Caspersson and Brandt's[6] findings concerning the yeast cell were based on the incomplete cytological information on the yeast cell available at that time.

An Hypothesis of the Rôle of Volutin (Metaphosphate) in Conferring Specificity on Cellular Enzymes.—Wiame[2] has presented convincing evidence supporting the view that volutin is metaphosphate. Assuming that his theory is correct, one may develop the following hypothesis:

Metaphosphate is synthesized from orthophosphate on the chromosomes and is transferred to the nucleolus. The synthesis of ribosenucleoprotein occurring in the nucleolus depends upon a supply of metaphosphate originating in the chromosome, and the specificity of cellular enzymes may be associated with this phenomenon, since genes are known to control the specificity of various cellular enzymes. Von Euler and Janssen[9] showed that the apoenzyme (protein) components of cellular enzymes are responsible for their specificity, and this important fact was recently confirmed by Spiegelman, Reiner and Morgan;[10] Caspersson and Schultz showed that the cellular ribosenucleoproteins are synthesized in the nucleolus and transferred from the nucleolus to the cytoplasm. It is, therefore, proposed that the specificity of enzymes whose synthesis is under genetic control depends upon specific substances carried with the polymerized metaphosphate into the nucleolus.

Orthophosphate cannot be used directly in the formation of ribosenucleoprotein, but must first pass through the metaphosphate stage. The energy-rich phosphate bond in metaphosphate furnishes the energy which is required for the synthesis of the ribosenucleoprotein in the nucleolus, as well as of the proteins of the chromosome itself. The synthesis of ribosenucleoprotein occurring in the nucleolus must initiate the production of specific enzymes. This suggests that the metaphosphates carry specificity-conferring groups with them from the chromosome to the nucleolus, which give specificity to the ribosenucleoproteins synthesized in the nucleolus.

* This work was supported by grants from Southern Illinois University, Anheuser-Busch, Inc., the U. S. Public Health Service and the American Cancer Society.

[1] Sylvén, Bengt, *Acta. Chir. Scand., Supplement*, **66**, 151 pp. (1941).

[2] Wiame, J. M., *Fed. Proc. Am. Soc. Expt. Biol.*, **6**, 302 (1947).

[3] Lindegren, C. C., *Mycologia*, **47**, 767–780 (1945).

[4] Lindegren, C. C., and Lindegren, Gertrude, *Cold Spring Harbor Symposia on Quantitative Biology*, **11**, 115–129 (1947).

[5] Rafalko, J. S., *Stain Technology*, **21**, 91–93 (1946).

[6] Caspersson, T., and Brandt, K., *Protoplasma*, **35**, 507–526 (1941).

[7] Lindegren, C. C., *Nature*, **159**, 63 (1947).

[8] Caspersson, T., and Schultz, Jack, PROC. NAT. ACAD. SCI., **26**, 507–515 (1940).

[9] v. Euler, H., and Janssen, B., *Z. physiol. Chem.*, **169**, 226 (1927).

[10] Spiegelman, S., Reiner, John M., and Morgan, Ida, *Arch. Bio.*, **13** (1), 113–125 (1947).

BOUNDARIES OF ULC SETS IN EUCLIDEAN n-SPACE

BY M. H. A. NEWMAN

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MANCHESTER, ENGLAND

Communicated by P. A. Smith, March 5, 1948

In this note an example is given of an $(n-1)$-manifold with non-vanishing fundamental group, embedded rectilienarly in Euclidean n-space, R^n, and bounding there a domain with all its homotopy groups zero (including π_1). This leads to an example of an open set in R^n which is ULC^∞, but whose boundary is not $1-LC$, thus settling in the negative a question raised by Eilenberg and Wilder.[1]

1. Let A and B be oriented circles with one common point, o, and let a space Z be formed by identifying the boundaries of two circular discs with the loops $A^{-3}(BA)^2$ and $B^{-5}(BA)^2$, respectively. More precisely, the boundary of disc 1 is divided into 7 equal arcs which are mapped onto A^{-1}, A^{-1}, A^{-1}, B, A, B, A, respectively, topologically save that both end-points of each arc are mapped onto o; and similarly for disc 2. All boundary points of the discs are then identified with their images. This 2-dimensional space is familiar as an example of a polyhedron which is acyclic[2] but has $\pi_1 \neq 0$. That $\pi_1(Z) \neq 0$ is clear, since the only relations connecting its generators A and B are

$$(BA)^2 = A^3 = B^5, \tag{1}$$

which are among those satisfied by the generators of the icosahedral group; but $H^1(Z) = 0$, since the adjunction of $AB = BA$ to (1) leads to $A = B = 1$. Again, if C_1^2 and C_2^2 are 2-chains on simplicial subdivisions of the two discs, with boundaries A and B, respectively, any 2-cycle on Z is of the form $mC_1^2 + nC_2^2$. Its boundary, written additively, is $m(-A + 2B) + n(2A - 3B)$, and hence can only be zero identically in A and B if $m = n = 0$. There are therefore no non-zero 2-cycles, and $H^2(Z) = 0$.

2. If $n \geq 5$, some simplicial subdivision Σ_0^n of the n-sphere S^n contains a simplicial subdivision K_0^2 of Z as a closed subcomplex.[3] Let Σ_0^n and K_0^2 be changed by a derivation[4] into Σ_1^n and K_1^2, respectively. If then $K_1^n = St(K_1^2)$ in Σ_1^n, each simplex of $K_1^n - K_1^2$ is uniquely expressible as the join $\sigma_a\sigma_b$, where $\sigma_a \in K_1^2$, $\sigma_b \in \Sigma_1^n - K_1^n$. Hence each point x of $|K_1^n - K_1^2|$ is uniquely expressible in the form $(1-\tau)x_a + \tau x_b$, where $x_a \in |\sigma_a|$, $x_b \in |\sigma_b|$,

$\sigma_a\sigma_b$ being the carrier simplex of x, and $0 < \tau < 1$. (Cf. Lefschetz, *Algebraic Topology*, p. 292.) We denote the number τ by $\tau(x)$.

Let X be the closed set of points

$$|K_1^2| \cup \{x|\tau(x) \leqq {}^1/_2\}.$$

in $|K_1^n|$. It is evidently a polyhedron. We denote its boundary by M, and $S^n - X$ by Y. The set X is retractible, on itself, into $|K_1^2|$, by projection along the segments x_bx_a. Hence X and $X - M$ are acyclic; and therefore Y and $IntY$[5] are acyclic, by the Alexander Duality Theorem. Any semilinear mapping ϕ_0 of S^1 into Y can be extended to a semilinear mapping ϕ_1 of E^2 into S^n which, since $n \geq 5$, can by arbitrarily small displacements of the vertices be made into a mapping ϕ_2 of E^2 into $S^n - |K_1^2|$. By projection of ϕ_2 outwards along the segments x_ax_b we obtain a mapping of E^2 into Y which is an extension of ϕ_0. Hence $\pi_1(Y) = 0$. Similarly, $\pi_1(IntY) = 0$.

However, $\pi_1(M) \neq 0$. Let ψ_0 be a semilinear and topological mapping of S^1 onto the circle A in $|K_1^2|$, covering the 1-simplexes $|\sigma_i^1|$ $(i = 1, 2, \ldots, k)$ in that order. Let $\sigma_i^1 = x_ix_{i+1}$, $(x_{k+1} = x_1)$. For each i choose an n-simplex $\sigma_i^n = \sigma_i^1\sigma_i^{n-2}$ of K_1^n, such that $\sigma_i^{n-2} \epsilon\, \Sigma_1^n - K_1^n$. Let ϵ be a positive number such that for all i, $U(x_i, \epsilon) \subset X$. For each i let y_iz_i be a segment parallel to x_ix_{i+1} and similarly sensed, lying in the interior of $|\sigma_i^n|$ save for its end-points y_i, z_i, which are on the boundary and within ϵ of x_i and x_{i+1}, respectively. Join z_{i-1} to y_i by a path s_i in $(S^n - |K_1^2|) \cap U(x_i, \epsilon)$. The segments y_iz_i and the paths s_i together determine a loop $\psi_1(S^1)$ in $X - |K_1^2|$, which is deformable into a mapping ψ_2 of S^1 into the circle A, by projecting points of y_iz_i along segments x_ax_b (as above), and each point of s_i linearly onto x_i. This deformation is in X, and since $\psi_2(S^1)$ covers each σ_i^1 once positively, $\psi_2 \simeq \psi_0$ in A. Now ψ_0 is not null-homotopic in X, since this would imply $\psi_0 \simeq 0$ in $|K_1^2|$ by retraction, contrary to the fact. Hence ψ_2 and ψ_1 are not null-homotopic in X. Since ψ_1 is in $X - |K_1^2|$ it can be projected out along the segments x_ax_b to give a loop $\psi_3(S^1)$ in M which is not null-homotopic in X, nor *a fortiori* in M.

3. In a suitable subdivision of Σ_1^n the sets M, X and Y are the loci of subcomplexes M^{n-1}, X^n, Y^n; and M^{n-1} is an $(n-1)$-manifold. Indeed this is true if in the definitions of M, X and Y the sphere Σ_1^n is replaced by any once derived manifold M_1^n, and K_1^2 by any closed subcomplex K_1. For each 1-simplex $|ab|$, with a in K_1 and b not, contains one point c of the set M. If the edges ab are "broken"[6] one by one in any order, taking the points c as new vertices, then in the subdivided manifold M_2^n the rectilinear simplexes determined by the new vertices form a subcomplex M^{n-1} whose locus is precisely M; and if X^n is $St(K_1)$ in M_2^n, $|\overline{X^n}| = X$ and $|\overline{M_2^n - X^n}| = Y$. Let c_0 be the new vertex on the edge a_0b_0. By projection from

a_0 the part of M in each simplex $|a_0b_0\sigma|$ of M_1^n is mapped (1, 1) onto $IntY \cap |b_0\sigma|$. We thus obtain a topological and semilinear mapping of $|St(c_0)$ in $M^{n-1}|$ onto a neighborhood of b_0 in the boundary of $|E^n|$, where $E^n = St(a_0b_0)$ in M_1^n. Since M_1^n is a once-derived n-manifold, E^n is an n-element, and its boundary is a sphere. It follows that M^{n-1} is an $(n-1)$-manifold.

4. Let $B(o, 1)$ stand for the "box" $|\xi_r| \leqq 1$ in R^n with center o and side 2, and let $Y(o, 1)$ denote a semilinear homeomorph of some subdivision of Y, having one (closed) boundary $(n-1)$-simplex $\overline{|\sigma_0^{n-1}|}$ in $\xi_1 = -1$, and all other points in $IntB(o, 1)$. Let $B(a, \lambda)$ denote, for any point $a = (\alpha_r)$, and any positive λ, the box $|\xi_r - \alpha_r| \leqq \lambda$, and let $Y(a, \lambda) = fY(o, 1)$, where f is the linear mapping that sends $B(o, 1)$ into $B(a, \lambda)$. Finally let e_i be the point $(2^{-i+2}, 1 + 2^{-i}, 0, 0, \ldots, 0)$ of R^n, $Y_i = Y(e_i, 2^{-i})$ and

$$F = \bigcup_{i=4}^{\infty} Y_i \cup B(o, 1).$$

The set IntF is ULC$^\infty$. This is equivalent to the assertion that $IntF$ is LC^∞ at all points of FrF.[7] If $e_\infty = (0, 1, 0, \ldots, 0)$, then at all points except e_∞, FrF has the character of an $(n-1)$-manifold embedded rectilinearly in R^n. Therefore $IntF$ is certainly LC^∞ at all points of FrF except possibly e_∞. The sets $B(o, 1)$ and Y_i are all acyclic, $Y_i \cap Y_j = 0$ for $i \neq j$, and $Y_i \cap B(o, 1)$ is an $(n-1)$-simplex. Therefore[8] the set

$$F_j = \bigcup_{i=4}^{j} Y_i \cup B(o, 1)$$

is acyclic for each j. Since any cycle on a compact subset of $IntF$ is necessarily contained in some F_j it follows that $IntF$ is acyclic. Similarly $\pi_1(Y_i) = \pi_1(B(o, 1)) = 0$, and therefore $\pi_1(F_j) = 0$. Since any mapping $f\colon S^1 \to IntF$ is necessarily into F_j for some j, $\pi_1(IntF) = 0$. By Hurewicz's theorem it follows that $\pi_r(IntF) = 0$ for $r \geq 1$.

Let g_j be the point $(0, 1 - 2^{-j}, 0, 0, \ldots, 0)$ of R^n, and

$$F_j' = \bigcup_{i=j+4}^{\infty} Y_i \cup B(g_j, 2^{-j}).$$

Then F_j' is the homeomorph of F under a linear mapping of the whole of R^n onto itself. Hence $IntF_j'$ is the homeomorph of $IntF$, and therefore all its homotopy groups vanish. Moreover

$IntF \cap U_j = IntF \cap U(e_\infty, 2^{-j}) \subseteq IntF_j' \subseteq U(e_\infty, 2^{2-j}) = IntF \cap U_{j-2}$. Hence every mapping of a sphere into $U_j \cap IntF$ is null-homotopic in $U_{j-2} \cap IntF$, i.e., $IntF$ is LC^∞ at e_∞.

By Theorem 4 of Eilenberg and Wilder (*loc. cit.*) F is LC^∞. However, *FrF is not 1 − LC*. For any fixed j let M_j be the boundary of Y_j, and

f_0: $S^1 \to M_j$ a semilinear mapping, not null-homotopic in M_j. By projection from a suitable point of the $(n-1)$-simplex $|\sigma_j^{n-1}|$ in which Y_j meets $B(o, 1)$, f_0 can be deformed, on M_j, into a mapping g_0 in $M_j - |\sigma_j^{n-1}|$, and therefore in FrF. If g_0 were extendable to a mapping g_1: $E^2 \to FrF$, then $g_1|g_1^{-1}(M_j)$ could be extended over $|\sigma_j^{n-1}|$ to give a mapping of E^2 into M_j. This would imply $f \simeq g_0 \simeq 0$, in M_j, contrary to hypothesis. Thus in every neighborhood of e_∞ there is a loop in FrF which is not null-homotopic in FrF: FrF is not $1 - LC$ at e_∞.

On the other hand, it follows from a theorem of Wilder[9] that since $IntF$ is ulc^∞ (in the sense of homology), FrF is lc^∞.

[1] Eilenberg, S., and Wilder, R. L., "Uniform Local Connectedness and Contractibility," *Am. J. Math.*, **64**, 613–622, Problem 2 (1942).

[2] I.e., connected, and q-acyclic for $q \geq 1$. All homology groups are formed with integral coefficients.

[3] Σ_0^n, K_0^2, etc., are abstract complexes. $|\sigma|$ denotes the (open) Euclidean simplex corresponding to the abstract simplex σ, and $|K| = \bigcup|\sigma|$ for $\sigma \epsilon K$. If $K_1 \subseteq K_2$, StK_1 (in K_2) is the set of all simplexes with a vertex in K_1.

It is in fact possible to embed K_0^2 as a subcomplex in S^4, but the corresponding Y is not aspherical.

[4] "Derived complex" = subdivision having one vertex for each simplex of the original complex, as in the barycentric division of the Euclidean complex.

[5] $IntX$ = interior of X, FrX = frontier of $X = X - IntX$.

[6] Breaking an edge ab in an abstract simplicial complex K is replacing each simplex $ab\sigma$ (the case $\sigma = 1$ included) by the two simplexes $ac\sigma$ and $cb\sigma$, where c is a new vertex, not in K.

[7] Cf. Eilenberg and Wilder, *loc. cit.* A set $X \subseteq S$ is $q - LC$ at any point a of S (not necessarily in X) if given $\epsilon > 0$ there exists $\eta(\epsilon, a) > 0$ such that every mapping of S^q into $X \cap U(a, \eta)$ is extendable to a mapping of E^{q+1} into $X \cap U(a, \epsilon)$. "LC^p at a" = "$q - LC$ at a for $q \leq p$," "LC^∞ at a" = "$q - LC$ at a for all q."

[8] Vietoris, L., "Homologiegruppen der Vereinigung zweier Komplexe," *Monatshefte Math. Phys.*, **37**, 159–162 (1930). The analogous theorem for the fundamental group, Seifert, H., "Konstruktion 3-dimensionaler Raüme," *Ber. Sächs. Akad.*, **83**, 26–66 (1931).

[9] Wilder, R. L., "Generalized Closed Manifolds in n-Space," *Ann. Math.*, **35**, 876–908 (1934).

CYCLOTOMIC POWER CHARACTERS AND TRINOMIAL EQUATIONS IN A FINITE FIELD

By H. S. Vandiver

Department of Applied Mathematics, University of Texas

Communicated March 1, 1948

In another article[1] employing extensions of methods originally due to Kummer, we obtained relations concerning power characters of units in a cyclotomic field defined by θ, where θ is a primitive mth root of unity,

m being composite. Our investigations were not complete, however, as they did not apply to the case where m is prime. In the present paper we shall develop these methods much further and obtain a result (Theorem I) which may not only be applied to power characters but also trinomial equations in a finite field as well as to other associated topics.

We define

$$\Psi_{a,\,b}(\theta) = \sum_{h} \theta^{-bh+(a+b)\ \mathrm{ind}\ (g^h+1)},$$

where g is a primitive root of the field $F(q^t)$ of order t, θ is primitive mth root of unity with $m = ln$ and l an odd prime, n any integer > 0 with $(l, n) = 1$, and h ranges over all values of the set $0, 1, 2, \ldots, q^t - 2$ except h' where $g^{h'} = -1$ in $F(q^t)$. If q is odd then $h' = (q^t - 1)/2$, and if q^t is even $h' = 0$. Also, $q^t - 1 \equiv 0 \pmod{m}$. It is known[2] that

$$\Psi_{a,\,b}(\theta)\Psi_{a,\,b}(\theta^{-1}) = q^t, \tag{2}$$

if $a \not\equiv 0$, $b \not\equiv 0$, $a + b \not\equiv 0 \pmod{m}$; and

$$\Psi_{a,\,0}(\theta)\Psi_{0,\,b}(\theta) = -1 \tag{3}$$

$$\Psi_{0,\,0}(\theta) = q^t - 2. \tag{3a}$$

In another paper[3] we defined operations with polynomials, using exponents in a commutative ring R with a unity element and also defined what we called formal exponential differentiation of polynomials or quotients of polynomials of this kind and used these concepts to obtain certain congruences in an algebraic field.

For the particular application of this idea that we shall need here, it is sufficient to define the formal exponential derivative with respect to x of a polynomial,

$$f(x) = \alpha_n x^n + \alpha_{n-1}x^{n-1} + \ldots + \alpha_0,$$

with the α's in the ring R and x an indeterminate over R, as

$$\frac{\bar{d}f(x)}{\bar{d}x} = \bar{D}_x f(x)$$

$$= n\alpha_n x^n + (n-1)\alpha_{n-1}x^{n-1} + \ldots + \alpha_1 x,$$

that is, we differentiate $f(x)$ in the ordinary way and then multiply by x If A and B are two polynomials in x then we also define the formal exponential derivative of A/B as

$$\frac{B\bar{D}_x(A) - A\bar{D}_x(B)}{B^2}.$$

Suppose that $\theta = \alpha\beta$, where $\alpha = e^{2i\pi/l}$, $\beta = e^{2i\pi\ n}$; and set

$$\Psi_{a,\,b}(x\beta) = \sum_h (x\beta)^{-bh+(a+b)\text{ ind }(g^h+1)}$$

where, now, x is an indeterminate over the ring of algebraic integers in an algebraic field defined by β. From this we immediately obtain

$$\frac{1}{\Psi(x_1\beta)}\bar{D}\Psi(x\beta) = \sum_h \frac{-bh + (a+b)\text{ ind }(g^h+1)(x\beta)^{-bh+(a+b)\text{ ind }(g^h+1)}}{\Psi(x\beta)}, \tag{5}$$

where, since all our differentiation is with respect to x, we write $\bar{D}$ for $\bar{D}_x$ and since, in the main, a and b are fixed during our investigation we also write

$$\Psi_{a,\,b}(x\beta) = \Psi(x\beta).$$

Also, as (5) resembles the operation of finding the ordinary derivative of a logarithm, we write for brevity and suggestiveness,

$$\bar{D}\log\Psi(x\beta) = \frac{1}{\Psi(x\beta)}\bar{D}\Psi(x\beta).$$

It follows easily by induction that, as was stated in (10) of a previous paper[3] (p. 25),

$$\bar{D}^{(n)}(AB) = \sum_{i=0}^{n}\binom{n}{i}\bar{D}^{(i)}(A)\bar{D}^{(n-i)}(B). \tag{6}$$

Applying this idea to (5) gives

$$\frac{\overline{d^r}\log\Psi(x\beta)}{\overline{dx^r}} = \frac{\bar{D}^r\Psi(x\beta)}{\Psi(x\beta)} + (r-1)\bar{D}^{\overline{r-1}}\Psi(x\beta)\bar{D}\left(\frac{1}{\Psi(x\beta)}\right) + \ldots + \bar{D}\Psi(x\beta)D^{\overline{r-1}}\left(\frac{1}{\Psi(x\beta)}\right).$$

The relation (2) may be written in the form

$$\Psi(\alpha\beta)\Psi(\alpha^{m-1}\beta^{m-1}) = q^t. \tag{8}$$

Now α is a primitive lth root of unity with l prime, and $(x^l-1)/(x-1)$ is irreducible in the field $K(\beta)$, since $(l, n) = 1$, consequently if x is an indeterminate, over $K(\beta)$,

$$\Psi(x\beta)\Psi(x^{m-1}\beta^{m-1}) = q^t + V(x)\frac{x^l-1}{x-1}, \tag{9}$$

where $V(x)$ is a polynomial in x with coefficients integers in the field $K(\beta)$.

We shall now prove the following lemma. If $c \geqq 1$ and j is any integer $\geqq 0$, and $s \not\equiv 0 \pmod{n}$, we have

$$\bar{D}^{(j)}\left[\sum_{i=0}^{cm-1}(x\beta^s)^i\right]_{x=1} \equiv (\text{mod } l). \tag{10}$$

For $s \equiv 0 \pmod{n}$ we have

$$\bar{D}^{(j)}\left[\sum_{i=0}^{cm-1} x^i\right]_{x=1} \equiv 0 \pmod{l}, \tag{10a}$$

for $j \not\equiv 0 \pmod{\overline{l-1}}$. To prove this, we note that the left-hand member of (10) can be written as

$$\frac{d^j}{dx^j}\left[((x\beta^s)^{cm}-1)\left(\frac{1}{x\beta^s-1}\right)\right]. \tag{11}$$

since $\beta = e^{2i\pi/n}$ and $m = ln$, it follows that for $x = 1$, $(x\beta^s)^{cm} - 1 = 0$, and $\bar{d}^k((x\beta^s)^{cm} - 1) = (cm)^k x^{cm} = (cln)^k \equiv 0 \pmod{l}$ Therefore the application of (6) to (11) gives for $x = 1$ every term $\equiv 0 \pmod{l}$, provided that $\bar{D}^{(u-t)}\left(\frac{1}{x\beta^s-1}\right)$ has a denominator prime to l in the field $K(\beta)$, but each denominator will consist of some power of $\beta^s - 1$ with $s \not\equiv 0 \pmod{n}$. Now it is known[4] that $\beta^s - 1$ will divide n and since $(l, n) = 1$ our proviso is satisfied and (10) is established. To prove (10a) we note that the left-hand member may be written in the form

$$\sum_{h=0}^{cm-1} h^j = \sum_{v,\,t}(vl+t)^j,\ t = 0, 1, \ldots, l-1;\ v = 0, 1, \ldots, cn-1.$$

Hence for $j \not\equiv 0$,

$$\sum_{h=0}^{cm-1} h^j \equiv cn \sum_{t=1}^{-1} t^j \pmod{l}, \tag{11a}$$

and the second factor on the right is $\equiv 0 \pmod{l}$ if $j \not\equiv 0 \pmod{\overline{l-1}}$. This completes the proof of the lemma.

From (9) we obtain,[5] taking $i < l - 1$,

$$\Psi(x^{m-1}\beta^{m-1}) = \frac{1+lnv}{\Psi(x\beta)} + \frac{V(x\beta)\,x^l - 1}{\Psi(x\beta)\,x - 1}, \tag{12}$$

$$\bar{D}^{(i)}[\Psi(x^{m-1}\beta^{m-1})] = \bar{D}^{(i)}\left(\frac{1+lnv}{\Psi(x\beta)}\right) + \bar{D}^{(i)}\left(\left(\frac{V(x)}{\Psi(x\beta)}\right)\left(\frac{x^l-1}{x-1}\right)\right). \tag{13}$$

Taking $c = 1$ in (11a) we note that every term in the expansion of the second term of the right-hand member of (13) is $\equiv 0 \pmod{l}$ for $x = 1$, since from (2), (3) and (3a), $\Psi(\beta)$ is prime to (l). This gives for $i < l - 1$,

$$[\bar{D}^{(i)}\Psi(x^{m-1}\beta^{m-1})]_{x=1} \equiv \bar{D}^{(i)}\left(\frac{1}{\Psi(x\beta)}\right) \pmod{l}, \tag{14}$$

where the right-hand member stands for

$$\bar{D}^{(i)}\left[\frac{1}{\Psi(x\beta)}\right]_{x=1}.$$

Also it may be shown that

$$[\bar{D}^{(i)}\Psi(x^{m-1}\beta^{m-1})]_{x=1} \equiv (-1)^i[\bar{D}^{(i)}\Psi(x\beta^{-1})]_{x=1} \pmod{l}, \quad (15)$$

by expanding both members of the equation and noting that for $x = 1$, $x^{m-1} = x$.

Application of (14) and (15) to (7) gives, modulo l, for $r < l$

$$[\bar{D}^r \log \Psi(x\beta)]_{x=1} \equiv \bar{D}^{(r)}(\Psi(\beta))\Psi(\beta^{-1}) + (r-1)(-1)\bar{D}^{(r-1)}\Psi(\beta)\cdot \bar{D}^{(1)}\Psi(\beta^{-1}) + \ldots + (-1)^{r-1}\bar{D}^{(1)}\Psi(\beta)\bar{D}^{(r-1)}\Psi(\beta^{-1}), \quad (16)$$

where the symbol $\bar{D}^{(i)}(\Psi(\beta))$ stands for

$$[\bar{D}_x^{(i)}(\Psi(x\beta))]_{x=1}.$$

Set

$$-bh + (a+b)\ \text{ind}\ (g^h + 1) = C_h \quad (17)$$

and

$$-bk + (a+b)\ \text{ind}\ (g^k + 1) = C_k, \quad (18)$$

by which, together with (16) and the definition (1), we obtain

$$[\bar{D}^{(r)} \log \Psi(x\beta)]_{x=1} = \sum_h \sum_k (C_h^r C_k^0 - (r-1)C_h^{r-1}\dot{C}_k + \ldots)(\beta)^{c_h - c_k} \pmod{l}. \quad (19)$$

or

$$[\bar{D}^{(r)} \log \Psi(x\beta)]_{x=1} \equiv \sum_h \sum_k C_h(C_h - C_k)^{r-1}\beta^{c_h - c_k} \pmod{l}, \quad (20)$$

h and k ranging independently over the values $0, 1, 2, \ldots, q^t - 2$, excepting $(q^t - 1)/2$ if q is odd and zero if $q = 2$ Put

$$g^{k-h} = g^{h'};\ \frac{g^k + 1}{g^h + 1} = g^{k'}.$$

Then the right-hand member of (20) may be written

$$[-b\ \text{ind}\ (g^{h'} - 1) + (a+b)\ \text{ind}\ (g^{k'} - 1) - a\ \text{ind}\ (g^{k'} - g^{h'})] \times (bk' - (a+b)h')^{r-1}\beta^{bk' - (a+b)h'}, \quad (21)$$

where h' and k' each range over $1, 2, \ldots, q^t - 2$ but $h' \neq k'$. We now break (21) down into three parts

$$\sum_h \sum_k - \beta^{bk-(a+b)h}(bk - (a+b)h)^{r-1}\ \text{ind}\ (g^h - 1) \quad \text{(I)}$$

$$\sum_h \sum_k (a+b)\beta^{bk-(a+b)h}(bk-(a+b)h)^{r-1} \text{ ind } (g^k-1) \qquad \text{(II)}$$

$$-\sum_h \sum_k a\beta^{bk-(a+b)h}(bk-(a+b)h)^{r-1} \text{ ind } (g^k-g^h). \qquad \text{(III)}$$

where we now re-employ the symbols h and k to range independently over $1, 2, \ldots, q^t-2$, but $h \neq k$. We see that (I) may be written as

$$b\sum_h \beta^{-ah}(-ah)^{r-1} \text{ in } (g^h-1) - b\sum_h \sum_k \beta^{bk-(a+b)h}(bk-(a+b)h)^{r-1} \text{ ind } (g^h-1), \qquad (22)$$

where now h and k each range independently over all values from 1 to q^t-2, inclusive, without exception. If $b \not\equiv 0 \pmod{n}$, then for each fixed value of h, the second term of (22) can be expanded, and when we carry out the summation with respect to k then each term of that expansion will be divisible by l (except the last one), by the lemma (10), (which could have been stated in the form $\sum_{h=0}^{q^t-2} h^i\beta^{sh} \equiv 0 \pmod{l}$ for $\beta^s \neq 1$ and $0^\circ = 1$), and (22) reduces to

$$(-a)^{r-1}b\sum_h \beta^{-ah}h^{r-1} \text{ ind } (g^h-1) + b\sum_h \beta^{-(a+b)h}(-(a+b)h)^{r-1} \text{ ind } (g^h-1). \qquad \text{(IV)}$$

If $a+b \not\equiv 0 \pmod{n}$, then, in a similar fashion, (II) can be reduced to

$$-(-a)^{r-1}(a+b)\sum_k \beta^{-ak}k^{r-1} \text{ ind } (g^k-1) - \sum_k (bk)^{r-1}(a+b)\beta^{bk} \text{ ind } (g^k-1). \qquad \text{(V)}$$

To reduce (III) introduce in it the terms corresponding to $h=0$ and $k=0$ with $h=k$, to offset which we add

$$a(-(a+b))^{r-1}\sum_h \beta^{-(a+b)h}h^{r-1} \text{ ind } (g^h-1) + ab^{r-1}\sum_h \beta^{bh}h^{r-1} \text{ ind } (g^h-1). \qquad \text{(VI)}$$

Then in the resulting expression in (III) we set h_1+k_1 for k and h_1 for h, and this gives

$$-a\sum_{h_1}\sum_{k_1}(h_1 + \text{ind } (g^{k_1}-1))(bk_1-ah_1)^{r-1}\beta^{bk_1-ah_1} \qquad \text{(VIa)}$$

$$h_1 = 0, 1, 2, \ldots, q^t-2; \quad k_1 = 1, 2, \ldots, q^t-2,$$

or

$$-a\sum_{h_1}\sum_{k_1}\beta^{bk_1-ah_1}h_1(bk_1-ah_1)^{r-1} - a\sum_{h_1}\sum_{k_1}\beta^{bk_1-ah_1}(bk_1-ah_1)^{r-1} \text{ ind } (g^{k_1}-1).$$

By (10) this reduces to zero modulo l, if $a \not\equiv 0 \pmod{n}$. Hence we add

the quantities given in (IV), (V), and (VI) which were the terms of the right-hand side of (20) and obtain for $r < l$,

$$\left[\frac{\bar{d}^{r-1}}{\bar{d}x^{r-1}}\left(\frac{\bar{D}\Psi(x\beta)}{\Psi(x\beta)}\right)\right]_{x=1} \equiv (-1)^r a^r \sum_k \beta^{-ak} A(k) + (-1)^{r+1}(a+b)^r \sum_k \beta^{-(a+b)k} A(k) - b^r \sum_k \beta^{bk} A(k) \pmod{l}, \quad (23)$$

where k ranges over the values 1, 2, ..., $q^t - 2$, and $A(k) = k^{r-1}$ ind $(g^k - 1)$. This gives

THEOREM I. *If a and b are integers such that $a \not\equiv 0 \pmod{n}$, $b \not\equiv 0 \pmod{n}$, $a + b \not\equiv 0 \pmod{n}$, $\beta = e^{2i\pi/n}$; $\alpha = e^{2i\tau/l}$; $ln = m$; $(n, l) = 1$; l an odd prime and*

$$\Psi(x\beta) = \sum_h (x\beta)^{-bh+(a+b)\text{ ind }(g^h+1)},$$

where h ranges over the integers 0, 1, ..., $q^t - 2$, excepting $(q^t - 1)/2$; q is an odd prime, g is a primitive root in the finite field of order q^t; with $q^t \equiv 1 \pmod{m}$. Then the congruence (23) holds where $r < l$, and the $\bar{d}$ symbols on the left indicate exponential differentiation as defined in the first part of the present article.

We now consider other conditions related to (23). If $r = l - 1$ and $n \neq 1$ then in reducing (I), (II), and (III) we find, using (10) and (10a), that we obtain (V) and (VI) without the restrictions $b \not\equiv 0 \pmod{n}$ and $a + b \not\equiv 0 \pmod{n}$. To obtain (VIb) we encounter, however, the term $a^{l-2}\sum_h h_1^{l-1}\beta^{-ah}$ in the development and this is not $\equiv 0 \pmod{l}$ when $a \equiv 0 \pmod{n}$. Hence, we have the result that (23) holds for $a \not\equiv 0 \pmod{n}$, $b \not\equiv 0 \pmod{m}$, $a + b \not\equiv 0 \pmod{m}$. If $n = 1$, then for $r = l - 1$ we have $\beta = 1$ and

$$\sum_{h_1} h_1^r = \sum_{h_1} h_1^r \beta^{-ah} \not\equiv 0 \pmod{l},$$

but we see, however, because of (10a) that (23) holds for $a \not\equiv 0 \pmod{m}$, $b \not\equiv 0 \pmod{m}$, $a + b \not\equiv 0 \pmod{m}$, if $r < l - 1$. These results enable us to apply (23) to obtain the expressions found for lth power character $((1 - \alpha\beta)/\mathfrak{p})$, in a previous paper,[1] where $\mathfrak{p}$ is a prime ideal in $K(\theta)$. For $n = 1$, $r < l - 1$ we derive from (23) a result which may be written as

$$\left[\frac{d^r \log \Psi(e^w)}{dw^r}\right]_{w=0} \equiv ((-1)^r a^r + (-1)^{r-1}(a+b)^r - b^r)\sum_h A(k),$$

and from this we easily obtain a relation due originally to Kummer[6] involving the power character $(\epsilon/\mathfrak{q})$ where $\epsilon = (1 - \alpha^r)/(1 - \alpha)$ and $\mathfrak{q}$ is a prime ideal in $K(\alpha)$. Applications to trinomial equations in a finite

field as well as other applications to power characters will be given elsewhere.

[1] *Am. J. Math.*, **47**, 140–147 (1925).
[2] Mitchell, H. H., *Trans. Am. Math. Soc.*, **17**, 167 (1916).
[3] Vandiver, H. S., these PROCEEDINGS, **28**, 25 (1942).
[4] Hilbert, D., *Gesammelte Abhandlungen*, Bd. I, Berlin, Springer (1932), pp. 199, 203.
[5] The derivations of the formulas (12)–(20) which follow resemble some of the steps in our previous paper,[1] but from then on the treatment is a bit different.
[6] Vandiver, H. S., *Ann. Math.*, **30**, 487 (1929), relation (1).

THE STABILITY OF DIFFERENTIAL EQUATIONS WITH PERIODIC COEFFICIENTS

BY SYLVAN WALLACH

DEPARTMENT OF MATHEMATICS, JOHNS HOPKINS UNIVERSITY

Communicated by F. D. Murnaghan, March 5, 1948

The following theorem will be proved:

If $f(t)$ is a real, continuous, non-constant function of period π which is such that

$$n^2 \leqq f(t) \leqq (n+1)^2, \qquad (-\infty < t < \infty), \tag{1}$$

for some integer n, then the characteristic exponents of

$$x'' + f(t)x = 0 \tag{2}$$

are distinct and of stable type.

A weakened form of this theorem, a form in which the signs of equality are excluded in (1) for every t, must be known for a long time (see, for example, Borg,[1] where reference is made to a lecture of Breuling). Actually, the truth of the theorem, at least of the weaker wording, is indicated by the diagrams available for such cases as the equations of Mathieu and Meissner; cf., e.g., Strutt.[2] However, I could not find a proof for even the weaker form of the theorem in the literature.

Whether or not the periodic function $f(t)$ satisfies (1), there exist a number $\rho \neq 0$ and a solution $x = x(t) \not\equiv 0$ of (2) such that

$$x(t+\pi) = \rho x(t), \tag{3}$$

where ρ and $x(t)$ can be complex-valued. The number ρ is a root of the characteristic equation of (2), a reciprocal quadratic equation with real coefficients. Hence, if ρ is not real, its absolute value is 1. If the trivial case where $f(t)$ is a constant is excluded, the assertion to be proved is that ρ is not real in virtue of (1).

Suppose, if possible, that ρ is real. Then if $\phi(t)$ is defined, in terms of the solution $x(t)$ occurring in (3), by

$$x(t) = |\rho|^{t/\pi}\phi(t),$$

it is seen that

$$\phi(t + \pi) = \epsilon\,\phi(t), \text{ where } \epsilon = sgn\rho = \pm 1.$$

Hence $\phi(t)$ has 2π as a period. Since $f(t)$ is real, the real and imaginary parts of $x(t)$ are solutions of (2). Consequently, (2) possesses a non-trivial solution of the form

$$x(t) = |\rho|^{t/\pi}\psi(t) \not\equiv 0, \tag{4}$$

where $\psi(t)$ is real and of period 2π.

Since the solution (4) is real, it follows from the inequalities (1), from Sturm's comparison theorem, and from the assumption that $f(t)$ is not constant, that

$$2n < N < 2(n + 1), \tag{5}$$

where N denotes the number of zeros of the solution (4) on a half-open interval of length 2π, say on $0 \leqq t < 2\pi$.

On the other hand, (4) shows that $x(t)$ and $\psi(t)$ have the same zeros. Since $\psi(t)$ is periodic, the number N of zeros of $\psi(t)$ on a period, $0 \leqq t < 2\pi$, is even. Since this contradicts (5), the proof is complete.

[1] Borg, G., *Ark. f. Matemat., Astr. o. Fysik*, **31**, No. 1, p. 28.

[2] Strutt, M. J. O., *Lamésche, Mathieusche und Verwandte Funktionen in Physik und Technik*, Berlin, 1932, pp. 24 and 40.

ON SOME EXPONENTIAL SUMS

By André Weil

Department of Mathematics, University of Chicago

Communicated by Marshall Stone, March 9, 1948

It seems to have been known for some time[1] that there is a connection between various types of exponential sums, occurring in number-theory, and the so-called Riemann hypothesis in function-fields. However, as I was unable to find in the literature a precise statement for this relationship, I shall indicate it here, and derive from it precise estimates for such sums, including the Kloosterman sums.

Let k be a finite field of q elements; consider the field $k(t)$ of rational functions in one transcendental element t, with coefficients in k; geo-

metrically, this is the function-field, over the ground-field k, of a projective straight line. On that straight line, we consider divisors, i.e., formal sums of points with integral (positive or negative) coefficients; and we limit ourselves, once for all, to divisors which are rational over k, i.e., such that conjugate points over k have the same coefficient. A divisor is called finite if it does not contain the point at infinity with a non-zero coefficient. Except for the notation, finite positive divisors are essentially the same as ideals in the ring $k[t]$; to every such divisor $\mathfrak{a}$, we attach the polynomial $P_{\mathfrak{a}}(t) = t^n + a_1 t^{n-1} + \ldots + a_n$ which generates the corresponding ideal, i.e., whose zeros are the points in $\mathfrak{a}$, with multiplicities respectively equal to their coefficients in $\mathfrak{a}$; as $\mathfrak{a}$ is assumed to be rational over k, $P_{\mathfrak{a}}(t)$ has its coefficients in k; and n is the degree of $\mathfrak{a}$. Every finite divisor $\mathfrak{m}$ can be written as $\mathfrak{m} = \mathfrak{a} - \mathfrak{b}$, where $\mathfrak{a}$, $\mathfrak{b}$ are finite positive divisors; to $\mathfrak{m}$, we attach the function $R_{\mathfrak{m}}(t) = P_{\mathfrak{a}}(t)/P_{\mathfrak{b}}(t)$; we have $\mathfrak{m} \sim 0$ if and only if $\mathfrak{a}$ and $\mathfrak{b}$, i.e., $P_{\mathfrak{a}}(t)$ and $P_{\mathfrak{b}}(t)$, are of the same degree, and then there is one and only one function in $k(t)$ having $\mathfrak{m}$ as its divisor and taking the value 1 at infinity, viz., $R_{\mathfrak{m}}(t)$ itself.

Let χ be a character of the multiplicative group k^* of the non-zero elements in k. Let $\mathfrak{d}$ be a finite divisor, consisting of the points ξ_ν with the coefficients a_ν; if $R(t)$ is in $k(t)$, we shall write $R(\mathfrak{d}) = \Pi_\nu R(\xi_\nu)^{a_\nu}$ whenever none of the $R(\xi_\nu)$ is 0 or ∞; as $\mathfrak{d}$ is rational over k, $R(\mathfrak{d})$ is in k. We shall assume that no a_ν is a multiple of the order of χ.

Furthermore, let ω be a character of the multiplicative group of power series in an indeterminate T with coefficients in k; we assume that ω has the value 1 for every series reduced to a monomial cT^m. According to the usual definition, we say that ω has the conductor (T^N) if it has the value 1 for every power-series which is $\equiv 1$ mod. T^N, and if N is the smallest integer with that property. Then the values of ω are p^s-th roots of unity, if p is the characteristic of k, and s is such that $p^s \geqslant N$. We shall write, for $x \in k$, $\lambda(x) = \omega(1 - xT)$. To every function $R(t)$ in $k(t)$, we can attach a power-series $R(1/T)$, arising from the expansion of the rational function $R(1/T)$ according to increasing powers of T; this is no other than the usual expansion of $R(t)$ at infinity. Then $\omega[R(1/T)]$ is defined; in particular, as $\omega(T) = 1$, we have $\omega[(1 - xT)/T] = \lambda(x)$. Now, for every finite divisor $\mathfrak{m}$ with no point in common with $\mathfrak{d}$, we write

$$\varphi(\mathfrak{m}) = \omega[R_{\mathfrak{m}}(1/T)] \cdot \chi[R_{\mathfrak{m}}(\mathfrak{d})]. \qquad (1)$$

This depends multiplicatively upon $\mathfrak{m}$, i.e., $\varphi(\mathfrak{m} + \mathfrak{n}) = \varphi(\mathfrak{m})\varphi(\mathfrak{n})$. Furthermore, if $\mathfrak{m} \sim 0$, and if there is a function $R(t)$ in $k(t)$, having $\mathfrak{m}$ as its divisor, taking the value 1 at every point ξ_ν, and such that $R(1/T) \equiv 1$ mod. T^N, we have $\varphi(\mathfrak{m}) = 1$; for $R(t)$ can then be no other than $R_{\mathfrak{m}}(t)$. According to well-known definitions, this shows that $\varphi(\mathfrak{m})$ is an Abelian character over the field $k(t)$, whose conductor consists of the point at

infinity with the coefficient N, and of the points ξ_ν with the coefficient 1; if d is the number of points ξ_ν in $\mathfrak{b}$, the degree of that conductor is therefore $N + d$; hence, by a known theorem,[2] the L-series belonging to this character is a polynomial of degree $N + d - 2$; calling α_i its roots, we have thus

$$\sum_{\mathfrak{a}} \varphi(\mathfrak{a}) \cdot u^{n(\mathfrak{a})} = \prod_{i=1}^{N+d-2} (1 - \alpha_i u), \tag{2}$$

where the sum in the left-hand side is extended to all finite positive divisors $\mathfrak{a}$ with no point in common with $\mathfrak{b}$, and where $n(\mathfrak{a})$ is the degree of $\mathfrak{a}$. Writing that the terms in u are equal on both sides, we get

$$\sum \varphi(\mathfrak{a}) = -\sum_i \alpha_i, \tag{3}$$

where the sum in the left-hand side is now extended only to the finite positive divisors of degree 1. These are in one-to-one correspondence with the polynomials $P_a(t) = t - x$, with $x \in k$. For such a divisor, we have

$$R_{\mathfrak{a}}(1/T) = P_a(1/T) = (1 - xT)/T,$$

hence $\omega[R_{\mathfrak{a}}(1/T)] = \lambda(x)$, and also

$$R_{\mathfrak{a}}(\mathfrak{b}) = \prod_\nu (\xi_\nu - x)^{a_\nu} = (-1)^a R_{\mathfrak{b}}(x),$$

with $a = \Sigma a_\nu$. Then (3) can be written as

$$\sum \lambda(x)\chi[R_{\mathfrak{b}}(x)] = (-1)^{a+1} \sum \alpha_i, \tag{4}$$

where the sum in the left-hand side is over all the elements x of k, other than the ξ_ν if any of these is in k. We may extend that sum to all elements x of k by agreeing that $\chi(0) = \chi(\infty) = 0$.

By class-field theory, the character $\varphi(\mathfrak{m})$ belongs to an Abelian extension of $k(t)$, and its L-series divides the zeta-function of that extension. Therefore, by the Riemann hypothesis,[3] all the α_i have the absolute value $\sqrt{q}$, hence

$$|\sum \lambda(x)\chi[R_{\mathfrak{b}}(x)]| \leqslant (N + d - 2)\sqrt{q}. \tag{5}$$

For instance, we can define a character ω, of conductor (T^2), by putting, for every series of constant term 1:

$$\omega(1 + x_1 T + x_2 T^2 + \ldots) = -\psi(x_1),$$

where ψ is a character of the additive group of k, not everywhere equal to 1. This gives

$$|\sum \psi(x)\chi[R_{\mathfrak{b}}(x)]| \leqslant d\sqrt{q}.$$

If the characteristic p of k is not 2, we have $d = 2$ for $R_{\mathfrak{b}}(t) = t^2 - a$, $a \neq 0$; hence, in that case,

$$|\sum\psi(x)\chi(x^2 - a)| \leqslant 2\sqrt{q}.$$

If, in this, we take for χ the character of k^* of order 2 (equal to 1 for squares, and to -1 for non-squares, in k^*), an elementary transformation[4] shows that the sum in the left-hand side is identical with the so-called Kloosterman sum $\Sigma\psi(cx + dx^{-1})$, for $4cd = a$; hence

$$|\sum_{x\neq 0}\psi(cx + dx^{-1})| \leqslant 2\sqrt{q},$$

and, in the case of a prime field of p elements, with $p \neq 2$:

$$|\sum_{x=1}^{p-1} e^{2\pi i/p\,(cx + d/x)}| \leqslant 2\sqrt{p}.$$

Furthermore, it is easily seen, e.g., by induction on n, that, if $F(x)$ is[1] any polynomial in x of degree n, with coefficients in k, such that $F(0) = 0$, there exists at least one character ω, of conductor (T^N) for some $N \leqslant n + 1$, such that, with the above notations, $\lambda(x) = \psi[F(x)]$. Then (5) gives:

$$|\sum\psi[F(x)]\chi[R_b(x)]| \leqslant (n + d - 1)\sqrt{q}.$$

[1] Cf., for example, H. Rademacher's excellent report on analytic number theory, *Bull. A. M. S.*, **48**, 379–401 (1942).

[2] Weissinger, J., *Hamb. Abhandl.*, **12**, 115–126 (1938).

[3] Weil, A., *Pub. Inst. Math. Strasbourg* (*N.S.*, no. 2), pp. 1–85 (1948).

[4] Davenport, H., *Crelles J.*, **169**, 158–176 (1933); cf. in particular Th. 5, p. 172.

ON SPACES WITH VANISHING LOW-DIMENSIONAL HOMOTOPY GROUPS

BY GEORGE W. WHITEHEAD

DEPARTMENT OF MATHEMATICS, BROWN UNIVERSITY

Communicated by S. Lefschetz, February 17, 1948

This note contains an investigation of the relationships between some of the homotopy and homology groups of an $(n - 1)$-connected space (i.e., a pathwise connected topological space whose homotopy groups of dimensions $< n$ all vanish).

Let X be an $(n - 1)$-connected space, and let A be a set of generators for the nth homotopy group, $\pi_n(X)$. For each $\alpha \in A$, let E_α^{n+1} be an $(n + 1)$-cell with boundary S_α^n; let y_α be a fixed reference point of S_α^n; x_0 a fixed reference point in X; and let f_α: $(S_\alpha^n, y_\alpha) \to (X, x_0)$ be a mapping representing the element $\alpha \in \pi_n(X)$. Suppose that $\bigcup_{\alpha\in A} E_\alpha^{n+1}$ is topologized so that the cells E_α^{n+1} are mutually separated and let E be the topological space obtained from $\bigcup_{\alpha\in A} E_\alpha^{n+1}$ by identifying all the points y_α to a single

point y_0. Let S be the subset of E obtained from $\bigcup_{\alpha \in A} S_\alpha^n$ by the above identification. Since the mappings f_α carry y_α into x_0, they induce a mapping $f:(S, y_0) \to (X, x_0)$. Finally let X^* be the identification space obtained from $E \cup X$ by identifying each point $y \in S$ with its image $f(y) \in X$. Then X may, in an obvious way, be considered as a subspace of X^* and the above identification induces a map $F:(E, S) \to (X^*, X)$ such that $F|S = f$. It is easy to see that X^* is n-connected. The results of this note are based on a consideration of the relations between the homology and homotopy groups of X^* and X.

The map $F:(E, S) \to (X^*, X)$ induces homomorphisms $F_k:\pi_{n+k}(E, S) \to \pi_{n+k}(X^*, X)$. The identity map $p^\alpha:(E_\alpha^{n+1}, S_\alpha^n) \to (E, S)$ induces isomorphisms $p_k^\alpha:\pi_{n+k}(E_\alpha^{n+1}, S_\alpha^n) \to \pi_{n+k}(E, S)$.

We shall make use of the following unpublished results of J. H. C. Whitehead:

If $k < n - 1$, then (1) *$\pi_{n+k}(E, S)$ is the weak direct sum of the subgroups $p_k^\alpha(\pi_{n+k}(E_\alpha^{n+1}, S_\alpha^n))$.* (2) *$F_k$ is an isomorphism onto.* Hence for $k < n - 1$, $\pi_{n+k}(X^*, X) \approx \sum_{\alpha \in A} \pi_{n+k}(E_\alpha^{n+1}, S_\alpha^n) \approx \sum_{\alpha \in A} \pi_{n+k-1}(S_\alpha^n)$.

The group $\pi_{n+k+1}(E^{n+k}, S^{n+k-1})$ is isomorphic under the boundary operator with $\pi_{n+k}(S^{n+k-1})$; it is cyclic of order two if $n + k > 3$. Let $h_k:(E^{n+k+1}, S^{n+k}) \to (E^{n+k}, S^{n+k-1})$ represent a generator of this group, and let $h_k^0 = h_k|S^{n+k}$. By the Freudenthal "suspension" theorems we may assume that h_{k+1}^0 is the suspension of h_k^0.

If $n + k > 3$, superposition with h_k induces a homomorphism $\bar{\eta}_k:\pi_{n+k}(X^*, X) \to \pi_{n+k+1}(X^*, X)$; if $\alpha \in \pi_{n+k}(X, X)$ is represented by a mapping $g:(E^{n+k}, S^{n+k-1}) \to (X^*, X)$, then $\bar{\eta}_k(\alpha)$ is represented by $gh_k:(E^{n+k+1}, S^{n+k}) \to (X^*, X)$. Similarly, superposition with h_{k+1}^0 induces homomorphisms

$$\eta_k:\pi_{n+k}(X) \to \pi_{n+k+1}(X)$$
$$\eta_k^*:\pi_{n+k}(X^*) \to \pi_{n+k+1}(X^*).$$

The homomorphisms η_k, η_k^*, $\bar{\eta}_k$ combine with the homomorphisms of the homotopy and homology sequences of (X^*, X) and the natural homomorphisms of the homotopy sequence into the homology sequence to produce a double ladder:

$$\begin{array}{ccccccccccc}
 & & \partial_2 & & i_1 & & j_1 & & \partial_1 & & \\
\cdots & \to & \pi_{n+1}(X) & \to & \pi_{n+1}(X^*) & \to & \pi_{n+}(X^*, X) & \to & \pi_n(X) & &
\end{array}$$

$$\begin{array}{ccccccccccccc}
 & \partial_3 & \downarrow\eta_1 & i_2 & \downarrow\eta_1^* & j_2 & \downarrow\bar{\eta}_1 & \partial_2 & \downarrow\eta_0 & i_1 & & j_1 & \\
\cdots \to & \pi_{n+2}(X) & \to & \pi_{n+2}(X^*) & \to & \pi_{n+2}(X^*, X) & \to & \pi_{n+1}(X) & \to & \pi_{n+1}(X^*) & \to & & \\
 & & & & & & & & & & & \partial_1 & \\
 & & & & & & & & & & & \pi_{n+1}(X^*, X) & \to \pi_n(X)
\end{array}$$

$$\begin{array}{ccccccccccccc} & & \partial'_3 & \downarrow\rho_2 & i'_2 & \downarrow\rho_2^* & j'_2 & \downarrow\bar{\rho}_2 & & \partial'_2 & \downarrow\rho_1 & i'_1 & \downarrow\rho_1^* & j'_1 \\ \ldots \to & H_{n+2}(X) & \to & H_{n+2}(X^*) & \to & H_{n+2}(X^*, X) & \to & H_{n+1}(X) & \to & H_{n+1}(X^* & \to & & & \\ & & & & & & & & & & \downarrow\bar{\rho}_1 & & \partial'_1 & \downarrow\rho_0 & \\ & & & & & & & & & & H_{n+1}(X^*, X) & \to & H_n(X) & & \end{array}$$

The homomorphisms denoted by ρ are the natural homomorphisms of the homotopy groups into the homology groups; the homomorphisms i_k and i'_k are those induced by the identity map $i: X \to X^*$; the homomorphisms j_k and j'_k are those induced by the identity map $j: (X^*, \phi) \to (X^*, X)$; the homomorphisms ∂_k and ∂'_k are those induced by the homotopy and homology boundary operators. The rows of the diagram are the homotopy and homology sequences of the pair (X^*, X); they are each *exact*[2] in the sense that the kernel of each homomorphism is the image of the preceding one. The homomorphisms of the diagram satisfy commutativity relations which state that for each square in the diagram, the two homomorphisms of the group in the upper left hand corner of the square into the group in the lower right hand corner are the same.

In the arguments to follow use is made of the Hurewicz theorems[3] that the homomorphisms ρ_0, $\bar{\rho}_1$ and ρ_1^* are isomorphisms onto. Examination of the diagram then yields the following results: (1) If $n > 1$, ρ_1 is onto. This result was proved by Fox[2] and independently by Hopf[4] for the case where X is a complex. (2) If $n > 2$, the kernel of ρ_1 is the image of η_0. Hence we have:

THEOREM 1. *If $n > 2$, $H_{n+1}(X) \approx \pi_{n+1}(X)/Image\ \eta_0$.*

Suppose now that $n > 3$. Then $\bar{\eta}_1$ is onto and the kernel of $\bar{\eta}_1$ is $2\pi_{n+1}(X^*, X)$. Moreover, since $\pi_{n+1}(S^n)$ is cyclic of order two, η_0 maps $2\pi_n(X)$ into zero. Define a homomorphism $\varphi: \pi_{n+2}(X^*) \to \pi_n(X)/2\pi_n(X)$ as follows: if $\alpha \in \pi_{n+2}(X^*)$, choose $\beta \in \pi_{n+1}(X^*, X)$ such that $\bar{\eta}_1(\beta) = j_2(\alpha)$ and let $\varphi(\alpha)$ be the coset of $2\pi_n(X)$ which contains $\partial_1(\beta)$. Then the image of φ is Kernel $\eta_0/2\pi_n(X)$ and the kernel of φ is the subgroup Q of $\pi_{n+2}(X^*)$ generated by the subgroups Kernel j_2 and Image η_1^*. Since the homotopy sequence is exact, Kernel j_2 = Image i_2; and since X^* is n-connected and $n + 1 > 2$, Image η_1^* = Kernel ρ_2^*. It follows easily that $\pi_{n+2}(X^*)/Q \approx H_{n+2}(X)/\text{Image } \rho_2$.

The group Image ρ_2 is the group $\sum_{n+2}(X)$ of spherical homology classes. Hence we have:

THEOREM 2. *If $n > 3$, $H_{n+2}(X)/\sum_{n+2}(X) \approx$ Kernel $\eta_0/2\pi_n(X)$.*

Let M^{n+2} be the join of the complex projective plane M^4 with an $(n-3)$-sphere. The $(n+2)$-nd integral homology group of M^{n+2} is infinite cyclic; let z be one of its generators. It then follows easily that

COROLLARY. *If $n > 2$ and $w \in H_{n+2}(X)$, there is a map $f: M^{n+2} \to X$ such that $f_*(z) = w$.*

If we further assume that $\pi_{n+2}(X) = 0$, further results can be obtained.

In this case $H_{n+3}(X)/\sum_{n+3}(X)$ is isomorphic with $H_{n+3}(X^*)/\sum_{n+3}(X^*)$. The latter group can be computed by applying Theorem 2 to the space X^* and we have[5]

THEOREM 3. *If $n > 3$ and $\pi_{n+2}(X) = 0$, then*

$$H_{n+3}(X)/\sum_{n+3}(X) \approx H_{n+1}(X)/2H_{n+1}(X) + {}_2(\pi_n(X)),$$

where ${}_2(\pi_n(X))$ is the subgroup of elements of $\pi_n(X)$ of order 2.

Summarizing, we see that *if X is an $(n-1)$-connected space and $n > 3$, the groups $H_{n+1}(X)$ and $H_{n+2}(X)/\sum_{n+2}(X)$ are determined by the homotopy groups $\pi_n(X)$ and $\pi_{n+1}(X)$ and the homomorphism $\eta_0:\pi_n(X) \to \pi_{n+1}(X)$. If also $\pi_{n+2}(X) = 0$, then $H_{n+2}(X)$ and $H_{n+3}(X)/\sum_{n+3}(X)$ are determined by $\pi_n(X)$, $\pi_{n+1}(X)$, and η_0.*

The above results are closely connected with recent work of Eilenberg and MacLane.[6] These authors have constructed for each abelian group π and each integer n an abstract complex $K(\pi, n)$ with the property that if X is an $(n-1)$-connected topological space such that $\pi_i(X) = 0$ for $n < i < n + h$, then the homology groups of dimension $< n + h$ of X and $K(\pi_n(X), n)$ are isomorphic, while the $(n + h)$th homology group of $K(\pi_n(X), n)$ is isomorphic with $H_{n+h}(X)/\sum_{n+h}(X)$. The kth homology group of $K(\pi, n)$ is denoted by $H_k(\pi, n)$, and the kth cohomology group of $K(\pi, n)$ with coefficients in an abelian group G is denoted by $H^k(\pi, n; G)$. The results above enable us to calculate explicitly some of the groups $H_k(\pi, n)$ with the following results:

THEOREM 4. *If $n > 1$, $H_{n+1}(\pi, n) = 0$. If $n > 3$, $H_{n+2}(\pi, n) \approx \pi/2\pi$. If $n > 3$, $H_{n+3}(\pi, n) \approx {}_2\pi$.*

In recent unpublished work,[7] Eilenberg and MacLane have defined for each space X such that $\pi_i(X) = 0$ for $i < n$ and $n < i < n + h$ a cohomology class k^{n+h+1} of $K(\pi_n(X), n)$ with coefficients in $\pi_{n+1}(X)$ and studied the influence of k^{n+h+1} on the homology groups of X. In the case $h = 1$ and $n > 2$, the cohomology class k^{n+2} will be described below.

Application of the universal coefficent theorem[8] shows that for any abelian group G the group $H^{n+2}(\pi, n; G)$ is naturally isomorphic with the group of homomorphisms of $H_{n+2}(\pi, n)$ into G. Call this isomorphism λ. Now if $n > 2$, η_0 is a homomorphism of $\pi_n(X)$ into $\pi_{n+1}(X)$ which maps $2\pi_n(X)$ into zero. Hence η_0 induces a homomorphism $\eta:\pi_n(X)/2\pi_n(X) \to \pi_{n+1}(X)$. Let μ be the isomorphism described above; $\mu:H_{n+2}(\pi_n(X), n) \approx \pi_n(X)/2\pi_n(X)$. Then $\eta\mu$ is a homomorphism of $H_{n+2}(\pi_n(X), n)$ into $\pi_{n+1}(X)$ and

THEOREM 5. $\lambda(k^{n+2}) = \eta\mu$.

The abstract complex $K(\pi, n)$ has enough simplicial structure that the $\cup_i$ products defined by Steenrod[9] for simplicial complexes can be introduced. As in Steenrod's paper, these products are used to define a homomorphism $Sq_{n-2}:H^n(\pi, n; G) \to H^{n+2}(\pi, n; G')$ whenever a commutative self-

pairing of G to G' is defined and the products $g_1 . g_2$ ($g_1, g_2 \in G$) all have order two.

The group $H^n(\pi, n; G)$ is isomorphic with the group of homomorphisms of $H_n(\pi, n)$ into G, and $H_n(\pi, n)$ is isomorphic with π. Hence $H^n(\pi, n; \pi)$ is isomorphic with the group of endomorphisms of π; let d^n be the element of $H^n(\pi, n; \pi)$ which acts as the identity endomorphism of π.

For $n > 2$, $\eta_0 : \pi_n(X) \to \pi_{n+1}(X)$ is a homomorphism with values in ${}_2(\pi_{n+1}(X))$ and therefore a commutative self-pairing of $\pi_n(X)$ to $\pi_{n+1}(X)$ can be found such that $\alpha . \alpha = \eta_0(\alpha)$. It then follows that $\alpha\ \beta$ always has order two and therefore $Sq_{n-2} : H^n(\pi_n(X), n; \pi_n(X)) \to H^{n+2}(\pi_n(X)$ $n; \pi_{n+1}(X))$ is defined and is independent of the particular choice of the self-pairing. Using this pairing, we find

THEOREM 6. *If* $n > 2$, $Sq_{n-2}(d^n) = k^{n+2}$.

[1] Freudenthal, H., *Comp. Math.*, **5**, 299–314 (1937).

[2] Fox, R. H., *Bull. Am. Math. Soc.*, **49**, abstract 172 (1943).

[3] Hurewicz, W., *Proc. Akad. Amsterdam*, **39**, 117–126 (1936).

[4] Hopf, H., *Comm. Math. Helv.*, **17**, 307–326 (1945).

[5] These results depend on Pontrjagin's result that $\pi_{n+2}(S^n) = 0$ for $n \geqq 3$. A proof was outlined in *C. R. Acad. Sci. U.R.S.S.*, **19**, 361–363 (1938), but a complete proof has not yet appeared.

[6] Eilenberg, S., and MacLane, S., *Ann. Math.*, **46**, 489–509 (1945).

[7] Cf. Eilenberg, S., and MacLane, S., *Proc. Nat. Acad. Sci.*, **32**, 277–280 (1946) for the case $n = 1$.

[8] Eilenberg, S., and MacLane, S., *Ann. Math.*, **43**, 757–831 (1942).

[9] Steenrod, N. E., *Ibid.*, **48**, 290–320 (1947).

GROUP THEORETICAL DISCUSSION OF RELATIVISTIC WAVE EQUATIONS

BY V. BARGMANN AND E. P. WIGNER

PRINCETON UNIVERSITY

Read before the Academy, November 18, 1947

Introduction.[1]—The wave functions, ψ, describing the possible states of a quantum mechanical system form a linear vector space **V** which, in general, is infinite dimensional and on which a positive definite inner product (ϕ, ψ) is defined for any two wave functions ϕ and ψ (i.e., they form a Hilbert space). The inner product usually involves an integration over the whole configuration or momentum space and, for particles of higher spin, a summation over the spin indices.

If the wave functions in question refer to a free particle and satisfy relativistic wave equations, there exists a correspondence between the wave functions describing the same state in different Lorentz frames.

The transformations considered here form the group of all *inhomogeneous* Lorentz transformations (including translations of the origin in space and time). Let $\psi_{l'}$ and ψ_l be the wave functions of the same state in two Lorentz frames l' and l, respectively. Then $\psi_{l'} = U(L)\psi_l$, where $U(L)$ is a linear unitary operator which depends on the Lorentz transformation L leading from l to l'. By a proper normalization, U is determined by L up to a factor ± 1. (For all details the reader is referred to the paper of reference 2, hereafter quoted as (L).) Moreover, the operators U form a single- or double-valued representation of the inhomogeneous Lorentz group, i.e., for a succession of two Lorentz transformations L_1, L_2, we have

$$U(L_2L_1) = \pm U(L_2)U(L_1). \tag{1}$$

Since all Lorentz frames are equivalent for the description of our system, it follows that, together with ψ, $U(L)\psi$ is also a possible state viewed from the original Lorentz frame l. Thus, the vector space $\mathbf{V}$ contains, with every ψ, all transforms $U(L)\psi$, where L is any Lorentz transformation.

The operators U may also replace the wave equation of the system. In our discussion, we use the wave functions in the "Heisenberg" representation, so that a given ψ represents the system for all times, and may be chosen as the "Schrödinger" wave function at time 0 in a given Lorentz frame l. To find ψ_{t_0}, the Schrödinger function at time t_0, one must therefore transform to a frame l' for which $t' = t - t_0$, while all other coordinates remain unchanged. Then $\psi_{t_0} = U(L)\psi$, where L is the transformation leading from l to l'.

A classification of all unitary representations of the Lorentz group, i.e., of all solutions of (1), amounts, therefore, to a classification of all possible relativistic wave equations. Such a classification has been carried out in (L). Two representations $U(L)$ and $\tilde{U}(L) = VU(L)V^{-1}$, where V is a fixed unitary operator, are equivalent. If the system is described by wave functions ψ, the description by

$$\tilde{\psi}_l = V\psi_l \tag{2}$$

is isomorphic with respect to linear superposition, to forming the inner product of two wave functions, and also to the transition from one Lorentz frame to another. In fact, if $\psi_{l'} = U(L)\psi_l$, then $\tilde{\psi}_{l'} = V\psi_{l'} = \tilde{U}(L)\tilde{\psi}_l$. Thus, one obtains classes of equivalent wave equations. Finally, it is sufficient to determine the irreducible representations since any other may be built up from them.

Two descriptions which are equivalent according to (2) may be quite different in appearance. The best known example is the description of the electromagnetic field by the field strength and the four vector potential, respectively. It cannot be claimed either that equivalence in the sense of (2) implies equivalence in every physical aspect. Thus, two equivalent

descriptions may lead to quite different expressions for the charge density or the energy density in configuration space (cf. Fierz,[3]) because (2) only implies global, but not local, equivalence of the wave functions. It should be emphasized, however, that any selection of one among the equivalent systems or the superposition of non-equivalent systems in any particular way involves an explicit or implicit assumption as to possible interactions, the positive character of densities, etc. Our analysis is necessarily restricted to free particles and does not lead to any assertions about possible interactions.

The present discussion is not based on any hypothesis about the structure of the wave equations provided that they be Lorentz invariant. In particular, it is not necessary to assume differential equations in configuration space. But it is a result of the analysis in (L) that every irreducible wave equation is equivalent (in the sense of (2)) to a system of differential equations. For the relation of the present point of view to other treatments of the subject see reference 11.

In the present note, we shall give, for every representation of (L), a differential equation the solutions of which transform according to that representation. We also will discuss in some detail the infinitesimal operators which generate the irreducible representations determined in (L), and we shall characterize these representations, and hence the covariant differential equations, by certain invariants constructed from the infinitesimal operators. This is of some interest, because the infinitesimal operators are closely related to dynamical variables of the system. L. Gårding[4] has recently shown that even in the infinite dimensional case one can rather freely operate with infinitesimal transformations. In particular, it immediately follows from his discussion (although it is not explicitly stated in his note) that the familiar commutation rules remain valid.

1. The Infinitesimal Operators of the Lorentz Group.—The metric tensor is assumed in the form $g_{44} = 1$, $g_{11} = g_{22} = g_{33} = -1$, $g_{kl} = 0$ $(k \neq l)$ and $g^{kl} = g_{kl}$. The scalar product of two four vectors a, b will be denoted by $\{a, b\} = a^k b_k$. Both c, the velocity of light, and $\hbar$, Planck's constant divided by 2π, are set equal to 1.

The Infinitesimal Operators p_k and M_{kl}. A translation in the x^k-direction is generated by p_k, a rotation in the $(x^k - x^l)$ plane by $M_{kl} = -M_{lk}$ $(k, l = 1, \ldots, 4)$. These operators are Hermitian, and the unitary operators U which represent the finite Lorentz transformation are obtained by exponentiation; thus $U = \exp(-i\alpha p_k)$ corresponds to a translation by the amount α in the direction x_k. Clearly, p_k are the four momenta of the system, and M_{23}, M_{31}, M_{12} the three components of the total angular momentum. The following commutation rules hold (where $[A, B] = AB - BA$)

$$[M_{kl}, M_{mn}] = i(g_{lm}M_{kn} - g_{km}M_{ln} + g_{kn}M_{lm} - g_{ln}M_{km}), \quad (3a)$$

$$[p_k, p_l] = 0 \qquad [M_{kl}, p_m] = i(g_{lm}p_k - g_{km}p_l). \quad (3b)$$

We now define four operators w_k by

$$(w^1, w^2, w^3, w^4) = (v_{234}, v_{314}, v_{124}, v_{321}), \quad (4a)$$

$$v_{klm} = p_k^* M_{lm} + p_l M_{mk} + p_m M_{kl} = M_{lm}p_k + M_{mk}p_l + M_{kl}p_m. \quad (4b)$$

Note that w_k is a "pseudo-vector," i.e., it is a vector only with respect to Lorentz transformations of determinant 1. By (3),

$$[M_{kl}, w_m] = i(g_{lm}w_k - g_{km}w_l) \qquad [p_k, w_l] = 0. \quad (5)$$

It follows from (3) and (5) that the two operators,

$$P = p^k p_k; \quad W = (1/6)v^{klm}v_{klm} = -w^k w_k, \quad (6)$$

commute with all the infinitesimal operators M_{kl} and p_k. Therefore, they have constant values (i.e., they are multiples of the unit operator) for every irreducible representation of the Lorentz group. (The familiar arguments which establish this for finite dimensional representations can be carried over to the infinite dimensional case. (Cf. V. Bargmann, reference 5, p. 602.))

W may also be written in the form

$$W = {}^1/_2 M_{kl}M^{kl}p_m p^m - M_{km}M^{lm}p^k p_l. \quad (7)$$

(This quantity was first introduced by W. Pauli, cf. Lubánski.[6]) The scalar product $w^k p_k$ vanishes.

2. *Summary of the Results of* (L).—(a) For every irreducible representation the states ψ may be expressed as functions $\psi(p, \xi)$ of the momentum vector p and an auxiliary variable ξ which may assume a finite or an infinite number of values. The momenta p are either all zero, or they vary over the manifold $p^k p_k = P$, with a constant value P. We confine ourselves to the cases in which $p \neq 0$, and either $P > 0$ or $P = 0$, because the remaining cases are unlikely to have direct physical significance.[5]

(b) To every inhomogeneous Lorentz transformation $y^k = \lambda^k{}_l x^l + a^k$ (in vector form: $y = \Lambda x + a$) corresponds a unitary operator $U(L)$ defined by

$$U(L)\psi(p, \xi) = e^{-i(a, p)}Q(p, \Lambda)\psi(\Lambda^{-1}p, \xi), \quad (8)$$

where $Q(p,\Lambda)$ is a unitary operator which may depend on p but affects only the variable ξ. The inner product (ϕ, ψ) is obtained by an integration over the manifold $p^k p_k = P$ and by a summation or integration over the variable ξ.

(c) The subgroup of the homogeneous Lorentz transformations which keep a fixed momentum vector p_0 unchanged is called "little group."

(The little groups defined by different vectors p_0 are isomorphic.) The unitary operators $Q(p_0, \Lambda)$ (where $\Lambda p_0 = p_0$) form an irreducible representation of the little group and determine the irreducible representation $U(L)$ of the inhomogeneous Lorentz group.

In all cases the operators M_{kl} have the form $\underset{0}{M}_{kl} + S_{kl}$, where the

$$\underset{0}{M}_{kl} = i\left(p_k \frac{\partial}{\partial p^l} - p_l \frac{\partial}{\partial p^k}\right) = i(p_k g_{lj} - p_l g_{kj}) \frac{\partial}{\partial p_j} \tag{9}$$

act on the variables p and correspond to the orbital angular momenta, while the S_{kl} act on the variables ξ and correspond to the spin angular momenta. Both $\underset{0}{M}_{kl}$ and S_{kl} satisfy the commutation rules (3a). Since the $\underset{0}{M}_{kl}$ do not contribute to v_{klm} (cf. (4b)), we have

$$v_{klm} = p_k S_{lm} + p_l S_{mk} + p_m S_{kl}; \quad [S_{kl}, p_m] = 0, \tag{10}$$

or, introducing the three-dimensional vector operators,

$$\vec{S} = (S_{23}, S_{31}, S_{12}); \quad \vec{S'} = (S_{14}, S_{24}, S_{34}); \quad \vec{p} = (p^1, p^2, p^3); \tag{10a}$$
$$\vec{w} = (w^1, w^2, w^3); \quad w^4 = \vec{p}.\vec{S}; \quad \vec{w} = p^4 \vec{S} - (\vec{p} \times \vec{S'}).$$

Clearly, M_{kl} may also be replaced by S_{kl} in the expression (7) for W.

For a fixed momentum vector p_0 the operators w_k are the infinitesimal generators of the little group. Since $w^k p_k = 0$, only three of them are linearly independent.

3. Classification of the Irreducible Representations.—We now turn to a brief summary of the main results, including the characterization of the representations in terms of the operators p and w. A more detailed discussion will follow in the succeeding sections.

The classes found in (L) (§§ 7, 8) are these:

I. P_s. *Particles of finite mass and spin s.*—Here $P = m^2 > 0$. In the rest system of the particle, the momentum vector has only the one non-vanishing component $p^4 = \pm m$, hence, by (10a), $W = m^2 S^2$. The operator $P^{-1}W$ represents the square of the spin angular momentum, and has the value $s(s+1)$ ($s = 0, 1/2, 1, \ldots$) for an irreducible representation. For a given momentum vector there are $2s + 1$ independent states. The representation $U(L)$ is single or double valued according to whether s is integral or half integral. The lowest cases ($s = 0, 1/2, 1$) correspond to the Klein-Gordon, Dirac and Proca equations, respectively.

II. 0_s. *Particles of zero rest mass and discrete spin.*—These representations may be considered limiting cases of the representations P_s for $m \to 0$. Then both P and W are equal to zero, and do not suffice to characterize these representations. For a given momentum vector, there exist 2 independent states if $s \neq 0$ (corresponding to two different states of polariza-

tion), and there is only one state if $s = 0$. Right and left circularly polarized states are described by the operator equations $w_k = sp_k$, and $w_k = -sp_k$, respectively, so that the representation 0_s is characterized by $P = 0$, $w_k w_l = s^2 p_k p_l$. The lowest cases ($s = 0, {}^1/_2, 1$) correspond to the scalar wave equation, the neutrino equation, and Maxwell's equations, respectively.

III. *$0(\Xi)$ and $0'(\Xi)$. Particles of zero rest mass and continuous spin.*—Here, $P = 0$, $W = \Xi^2$, where Ξ is a real positive number. For a given momentum vector there exist infinitely many different states of polarization, which may be described by a continuous variable. The representation $0(\Xi)$ is single valued, while $0'(\Xi)$ is double valued.

To construct these representations explicitly, we shall select, in each case, one among the equivalent sets of wave equations, define a Lorentz invariant inner product (ϕ, ψ), and prove the operator relations stated above. We shall operate in momentum space; this is particularly simple, because the momenta (but not the coördinates) are defined by the Lorentz group, as infinitesimal translations.

4. *The Class P_s.*—(*a*) $s = 0$. Here, the variable ξ assumes only one value and may therefore be omitted. Consequently, $Q(p, \Lambda) = 1$ (cf. reference 8), and for the little group the trivial one-dimensional representation is obtained. Hence, $S_{kl} = 0$, and $w_k = 0$. The wave equation reduces to $p^k p_k = m^2$; the inner product (ϕ, ψ) is determined by the norm (ψ, ψ) of a wave function,

$$(\psi, \psi) = \int |\psi(p)|^2 d\Omega, \text{ where } d\Omega = |p^4|^{-1} dp^1 dp^2 dp^3, \qquad (11)$$

the integral being extended over both sheets of the hyperboloid $p^k p_k = P = m^2$. The expression (11) is Lorentz invariant, because $d\Omega$ is an invariant volume element in momentum space. For the wave function in configuration space, one finds

$$\psi(x) = (2\pi)^{-3/2} \int e^{-i(p, x)} \psi(p) d\Omega, \qquad (12)$$

where x stands for x^1, x^2, x^3, x^4. It is well known that (ψ, ψ) cannot be simply expressed in configuration space, because for the Klein-Gordon equation the density is indefinite, and the integral over the density in configuration space coincides with (11) only if $\psi(p) = 0$ whenever $p^4 < 0$.

(*b*) $s = {}^1/_2 N$ with $N = 1, 2, 3, \ldots$. For particles of higher spin we use the equations first derived by Dirac[7] in the form essentially given in reference 8. We use for ξ the N four-valued variables $\zeta_1, \ldots, \zeta_N$ in which the wave function $\psi(p; \zeta_1, \ldots, \zeta_N)$ is symmetric. We define for every ζ_ν four-dimensional matrices γ_ν^k of the same nature as are used in Dirac's electron theory:

$$\gamma_\nu^k \gamma_\nu^l + \gamma_\nu^l \gamma_\nu^k = 2g^{kl} 1 \quad (k, l = 1, 2, 3, 4). \qquad (13)$$

The γ with different lower indices ν commute. The γ^1, γ^2, γ^3 are skew Hermitian, γ^4 is Hermitian. The wave equations then are

$$\gamma_\nu{}^k p_k \psi = m\psi \quad (\nu = 1, 2, \ldots, N). \tag{14}$$

It follows from any of these equations in well known fashion that

$$g^{kl} p_k p_l \psi = p^k p_k \psi = m^2 \psi. \tag{14a}$$

The infinitesimal operators of displacement are the p, those of four-dimensional rotation the $M_{kl} = \underset{0}{M}_{kl} + S_{kl}$ with $\underset{0}{M}_{kl}$ of (9) and

$$S_{kl} = {}^1/_2 i \sum_\nu \gamma_{\nu k} \gamma_{\nu l} \quad (k \neq l), \tag{15}$$

where the

$$\gamma_{\nu k} = g_{kl} \gamma_\nu{}^l \tag{15a}$$

satisfy the same relations (13) as do the $\gamma_\nu{}^k$.

The invariant scalar product is

$$(\psi, \psi) = \int \Big| \sum_\zeta \psi^* \gamma_1{}^4 \gamma_2{}^4 \ldots \gamma_N{}^4 \psi \Big| d\Omega. \tag{16}$$

In fact, (16) is invariant both with respect to the operators $\underset{0}{M}$ and also with respect to the S. The latter condition means that

$$((1 + i\epsilon S_{kl})\psi, (1 + i\epsilon S_{kl})\psi) = (\psi, \psi),$$

up to terms with ϵ^2. This formula can be verified by observing that, if both k and l are space like S_{kl} is a Hermitian matrix and commutes with the product of the γ^4. If either k or l is 4, S_{kl} is skew Hermitian, but anticommutes with the product of the γ^4. It follows that (16) is invariant with respect to the proper Lorentz transformations. Its invariance with respect to reflections, etc., can also be shown.

The absolute sign in (16) is necessary to make it positive definite. We now shall give (16) a new form which is based on the set of identities

$$(p_4)^\nu \gamma_\nu{}^4 \ldots \gamma_2{}^4 \gamma_1{}^4 \psi = m^\nu \psi + A_\nu \psi, \tag{17}$$

where A_ν is a skew Hermitian matrix involving only the first ν of the γ^k (and the p). We can prove (17) best by induction: applying $p_4 \gamma_{\nu+1}{}^4$ to (17) gives, by means of (14),

$$\begin{aligned}(p_4)^{\nu+1} \gamma_{\nu+1}{}^4 \gamma_\nu{}^4 \ldots \gamma_2{}^4 \gamma_1{}^4 \psi &= m^\nu p_4 \gamma_{\nu+1}{}^4 \psi + p_4 \gamma_{\nu+1}{}^4 A_\nu \psi \\ &= m^{\nu+1} \psi + (-m^\nu p_k \gamma_{\nu+1}{}^k + p_4 \gamma_{\nu+1}{}^4 A_\nu)\psi \quad (k = 1, 2, 3).\end{aligned} \tag{17a}$$

The last bracket is $A_{\nu+1}$: it is skew Hermitian and involves only the first $\nu + 1$ of the γ so that (17) is established by induction. Setting $\nu = N$ in (17) multiplying with ψ and summing over the ζ yields

$$p_4{}^N \sum_\zeta \psi^* \gamma_1{}^4 \gamma_2{}^4 \ldots \gamma_N{}^4 \psi = m^N \sum_\zeta |\psi|^2 + \sum_\zeta \psi^* A_N \psi. \quad (17b)$$

Because of the skew Hermitian nature of A_N, the last term is imaginary. Since the two other terms of (17b) are real, they must be equal. As a result, we can write for (16) also

$$(\psi, \psi) = \int |m/p_4|^N \sum_\zeta |\psi|^2 d\Omega. \quad (18)$$

At the same time, (18) permits us to give another form to the scalar product,

$$(\psi, \psi) = \int |p_4|^{-N-1} \sum_\zeta |\psi|^2 dp_1 dp_2 dp_3, \quad (18a)$$

which differs from (18) or (16) by the positive constant m^{-N}. It may be worth noting here that the absolute signs in (16), and in the definition (11) of $d\Omega$ (or in (18a)), can be omitted in case of an odd N. This makes it possible to define a simple positive definite scalar product in coördinate space by means of (12). In particular, for $N = 1$, (16) (or (18a)) equals the integral of $|\psi|^2$ over ordinary space. In case of even N (integer spin s) no simple positive definite scalar product can be defined in coördinate space.

It is now established that the solutions of (14) form a Lorentz invariant set in which a positive definite scalar product (16) or (18a) can be defined. We shall now determine the representation of §2 to which the solutions belong and will also calculate the invariants P and W.

In order to define a little group, we choose as momentum p_0 with the components 0, 0, 0, m. The little group then becomes the group of rotations in ordinary space. If we assume that the γ^4 are diagonal, with diagonal elements 1, 1, -1, -1, equation (14) shows that only those components of ψ can be different from zero which correspond to the first two rows of γ_ν. There are 2^N such components, the rest of the 4^N components of ψ must vanish. Even these components will not be independent: as a result of the symmetry of the ψ in the ζ, all components of ψ will be equal in which the same number κ of the N indices ζ correspond to the first row of the γ, the $N - \kappa$ other indices to the second row. Since κ can assume any of the values between 0 and N, there are $N + 1$ such components. If $p_4 = -m$, the same considerations will hold, except that the last two rows of γ will play the rôle which the first two rows play in case of $p_4 = m$.

In order to determine the transformation properties of these $N + 1 = 2s + 1$ independent components under the elements of the little group, we note that the space like M give zero if applied to ψ with a purely time like $p = p_0$. We need only to calculate, therefore, the effect of the S_{kl}

on ψ. Since, in particular, $^1/_2 i\gamma_1\gamma_2$ commutes with γ^4, but is not identical with it, we can assume that it is diagonal and has the diagonal elements $^1/_2$, $-^1/_2$, $^1/_2$, $-^1/_2$. If the sum of such $^1/_2 i\gamma_1\gamma_2$ is applied to the component of ψ in which κ of the ζ_ν correspond to the first row, $N - \kappa$ to the second row, this component will be multiplied by $^1/_2\kappa - ^1/_2(N - \kappa) = \kappa - s$. Since κ runs from 0 to $N = 2s$, the $M_{12}\psi = S_{12}\psi$ will run from $-s\psi$ to $s\psi$. Hence the representation of the little group in question is $D^{(s)}$, as was postulated.

Because of (10a), W becomes $m^2(S_{23}^2 + S_{31}^2 + S_{12}^2)$ or, since the S_{23}, S_{31}, S_{12} are the infinitesimal operators of $D^{(s)}$, we have $W = m^2 s(s + 1)$ as given[9] in §3. The value of P is m^2 because of (14a).

5. *The Class O_s.*—(*a*) $s = 0$. The corresponding discussion in the preceding section may be literally applied to this case, with the exception that $m = 0$ and that the integral (11) is to be extended over the light cone.

(*b*) The wave equations can be obtained by setting $m = 0$ in (14). The infinitesimal operators continue to be given by (9) and (15). The scalar product must be defined by (18a) because (16) vanishes for all ψ. The invariance of this scalar product follows from the invariance of (18a) for finite mass because, except at $p_1 = p_2 = p_3 = 0$, the wave function is continuous in m.

The essential difference between finite and zero mass is that, in the latter case, not only the infinitesimal operators but also the wave equation are invariant under any one of the operators $\Gamma_\nu = i\gamma_\nu{}^1\gamma_\nu{}^2\gamma_\nu{}^3\gamma_\nu{}^4$. As a result, for $m = 0$, the linear manifold defined by (14) can be decomposed into invariant manifolds by giving definite values to the Γ_ν. In particular, we shall be concerned henceforth with the manifold defined by (14) and

$$\Gamma_\nu\psi = \psi \qquad (\nu = 1, 2, \ldots, N), \tag{19a}$$

and with the other one for which

$$\Gamma_\nu\psi = -\psi \qquad (\nu = 1, 2, \ldots, N) \tag{19b}$$

holds. Both manifolds are invariant under proper Lorentz transformations but go over into each other by reflections: they correspond physically to right and left circular polarization.[10]

Let us now again choose a particular momentum vector p_0 in order to define the little group. The covariant components of p_0 shall be 0, 0, 1, 1. The wave equations (14) then can be written, after multiplication with $\gamma_\nu{}^3$, in the form

$$\gamma_\nu{}^3\gamma_\nu{}^4\psi = \psi \qquad (\nu = 1, 2, \ldots, N). \tag{20}$$

It is now advantageous to assume that the $\gamma^3\gamma^4$ are diagonal, their diagonal elements being 1, 1, -1, -1. Equation (20) then expresses the fact that ψ for the p_0 in question is different from zero only if all ζ have values corre-

sponding to the first two rows of the γ. Since the Γ commute with the $\gamma^4\gamma^3$ but are not identical with them, they may be also assumed to be diagonal, with diagonal elements $1, -1, 1, -1$. Hence, in the manifold defined by (20) and (19a) all components of ψ vanish (for $p = p_0$) unless all ζ have values corresponding to the first rows of the γ: the manifold (20), (19a) is one dimensional for given p. The same holds for the manifold defined by (20), (19b) except that in this case $\psi(p_0; \zeta_1, \ldots, \zeta_N)$ differs from zero only if all ζ have values corresponding to the second row of the γ. For given momentum, ψ has only two independent components.

The infinitesimal operators of the little group are M_{12}, $M_{13} - M_{14}$, $M_{23} - M_{24}$ which leave p_0 invariant. The corresponding M give again zero if applied to ψ at $p = p_0$. The S corresponding to the second of the above operators (cf. (15), (15a)) is a sum of matrices $^1/_2i(\gamma_\nu{}^1\gamma_\nu{}^3 + \gamma_\nu{}^1\gamma_\nu{}^4)$. It vanishes if applied to our ψ as can be seen by applying $\gamma_\nu{}^1\gamma_\nu{}^3$ to (20). The same holds for $M_{23} - M_{24}$. On the other hand, $^1/_2i\gamma_\nu{}^1\gamma_\nu{}^2$ gives $^1/_2\psi$ if applied to the ψ of (20), (19a), and gives $-^1/_2\psi$ if applied to the ψ of (20), (19b). One sees this most easily by applying $^1/_2i\gamma_\nu{}^1\gamma_\nu{}^2$ to (20) and making use of (19). As a result, $M_{12}\psi = \pm^1/_2N\psi = \pm s\psi$ for the two manifolds in question: these indeed belong to the representation 0_s of the inhomogeneous Lorentz group.

The value of the invariant P is zero. The above also involves a calculation of the w for the ψ at $p = p_0$: we have $w^3\psi = M_{12}\psi = \pm s\psi$, $w^1\psi = (M_{42} + M_{23})\psi = 0$, $w^2\psi = (M_{31} + M_{14})\psi = 0$, $-w^4\psi = M_{12}\psi = \pm s\psi$. It follows that the value of the second invariant $W = -(w^4)^2 + (w^1)^2 + (w^2)^2 + (w^3)^2$ is also zero for all the manifolds 0_s; these cannot be characterized by P and W. However, these manifolds can be characterized by the equation $P = 0$ with the additional set

$$w_k = sp_k \text{ and } w_k = -sp_k, \tag{21}$$

the $+$ applying to (19a), the $-$ to (19b). Both these equations are invariant with respect to proper Lorentz transformations. If reflections are to be included, one can combine them into $w_kw_l = s^2p_kp_l$.

6. The Class $0(\Xi)$.—Here, the auxiliary variable is a space like four vector ξ of length l, orthogonal to p. The scalar function $\psi(p, \xi)$ is determined by the equations[11]

$$g^{kl}p_kp_l\psi = 0;\ g^{kl}p_k\xi_l\psi = 0;\ g^{kl}\xi_k\xi_l\psi = -\psi, \tag{22}$$

$$p_k\partial\psi/\partial\xi_k = -i\Xi\psi, \tag{22a}$$

with a real positive constant Ξ. By (22a), for every real number ρ,

$$\psi(p, \xi + \rho p) = e^{-i\rho\Xi}\psi(p, \xi). \tag{23}$$

The infinitesimal operators of displacement are the p_k, those for rotations are the M of (9) plus the

$$S_{kl} = i\left(\xi_k \frac{\partial}{\partial \xi^l} - \xi_l \frac{\partial}{\partial \xi^k}\right) = i(\xi_k g_{lj} - \xi_l g_{kj}) \frac{\partial}{\partial \xi_j}. \tag{24}$$

In order to find the invariant scalar product, we introduce, for every vector p on the light cone, two real space like vectors $u^{(1)}(p)$ and $u^{(2)}(p)$ of length one, orthogonal to p and to each other, so that

$$\{u^{(r)}(p), p\} = 0, \ \{u^{(r)}(p), u^{(s)}(p)\} = -\delta_{rs} \quad (r, s = 1, 2). \tag{25}$$

Then ξ is a linear combination of p, $u^{(1)}(p)$, $u^{(2)}(p)$,

$$\xi = \alpha p + \beta_1 u^{(1)}(p) + \beta_2 u^{(2)}(p), \tag{26}$$

where α and the β are real. $\{\xi, \xi\} = -1$ implies $\beta_1^2 + \beta_2^2 = 1$, hence $\beta_1 + i\beta_2 = e^{i\tau}$ with a suitable real angle τ. $\psi(p, \xi)$ is therefore a function of p, α, τ,

$$\psi(p, \xi) = \phi(p, \alpha, \tau). \tag{27}$$

The choice of the $u^{(r)}(p)$ is, of course, not unique. Let $v^{(r)}(p)$ be another system of vectors which satisfy (25). They may be expressed in the form (26), i.e.,

$$v^{(r)}(p) = \kappa_r p + \sum_s \lambda_{rs} u^{(s)}(p) \quad (r, s = 1, 2).$$

By (25), the matrix λ_{rs} is orthogonal. In terms of the $v^{(r)}$, $\xi = \alpha' p + \sum_r \beta'_r v^{(r)}(p)$, where $\beta'_r = \sum_s \lambda_{rs}\beta_s$. In particular

$$\beta'_1 + i\beta'_2 = e^{i\tau'}; \quad \tau' = \pm(\tau + \lambda) \tag{28}$$

λ depending on the λ_{rs}. By (23), $|\phi(p, \alpha, \tau)| = |\phi(p, 0, \tau)|$, and we define the norm of ψ by

$$(\psi, \psi) = \int |\phi(p, 0, \tau)|^2 d\Omega d\tau. \tag{29}$$

This expression is independent of the choice of the $u^{(r)}$. In fact, let $\phi(p, \alpha, \tau) = \phi'(p, \alpha', \tau')$ where the primed variables refer to another set $v^{(r)}$. Then $|\phi(p, 0, \tau)| = |\phi'(p, \alpha', \tau')| = |\phi(p, 0, \tau')|$, and $|d\tau'/d\tau| = 1$. To prove the Lorentz invariance of (29) we proceed as follows: If a homogeneous Lorentz transformation maps p on $\Lambda^{-1}p$, and ξ on $\Lambda^{-1}\xi$, we may, in particular, choose the $u^{(r)}(p)$ in the new system to be the transforms of the original ones; then the coefficients α, β_1, β_2 in (26), and hence τ remain unchanged, and the integral (29) is invariant.

If we choose as the basic vector again p_0 with the components 0, 0, 1, 1 the infinitesimal operators of the little group are again M_{12}, $M_{13} - M_{14}$.

$M_{23} - M_{24}$. The M parts of these give zero for $p = p_0$, the S parts of the latter two are

$$S_{13} - S_{14} = -i\xi_1\left(\frac{\partial}{\partial\xi_3} + \frac{\partial}{\partial\xi_4}\right) + i(\xi_3 - \xi_4)\frac{\partial}{\partial\xi_1}, \qquad (30a)$$

$$S_{23} - S_{24} = -i\xi_2\left(\frac{\partial}{\partial\xi_3} + \frac{\partial}{\partial\xi_4}\right) + i(\xi_3 - \xi_4)\frac{\partial}{\partial\xi_2}. \qquad (30b)$$

Because of (22a), the first term gives, if applied to ψ at $p = p_0$ just $\Xi\xi_1\psi$ and $\Xi\xi_2\psi$, respectively. The second terms vanish because of the second equation of (22). Hence ψ is not invariant under the "displacements" $M_{13} - M_{14}$ and $M_{23} - M_{24}$ in ξ space, and the sum of the squares of the "momenta" is $(\xi_1^2 + \xi_2^2)\,\Xi^2 = \Xi^2$ because of the last equation of (22). This is also the value of W, while $P = 0$.

7. *The class* $0'$ (Ξ).—Since the discussion of this last case follows the pattern of the preceding section we confine ourselves to stating the main results. We introduce, in addition to the vector ξ, a discrete spin variable ζ which can assume four values. The wave equations become

$$\gamma^k p_k\psi = 0; \quad g^{kl}p_k\xi_l\psi = 0; \quad g^{kl}\xi_k\xi_l\psi = -\psi. \qquad (31)$$

$$p_k\partial\psi/\partial\xi_k = -i\Xi\psi \qquad (31a)$$

The parameters α and τ are introduced as before. The norm is given by

$$(\psi, \psi) = \int p_4^{-2}\sum|\phi(p, 0, \tau)|^2 dp_1 dp_2 dp_3 d\tau. \qquad (32)$$

(Cf. (18a) and (29).) Again $W\psi = \Xi^2\psi$, $P\psi = 0$.

It may be remarked that the scalar product has a simple positive definite form in coördinate space for these equations.[11]

[1] All the essential results of the present paper were obtained by the two authors independently, but they decided to publish them jointly.

[2] Wigner, E. P., *Ann. Math.*, **40**, 149–204 (1939).

[3] Fierz, M., *Helv. Phys. Acta*, **XII**, 3–37 (1939).

[4] Gårding, L., *Proc. Nat. Acad. Sci.*, **33**, 331–332 (1947).

[5] Gelfand, L., and Neumark, M., *J. Phys.* (*USSR*), **X**, 93–94 (1946); Harish-Chandra, *Proc. Roy. Soc.* (London), **A, 189**, 372–401 (1947); and Bargmann, V., *Ann. Math.*, **48**, 568–640 (1947), have determined the representations of the homogeneous Lorentz group. These are representations also of the inhomogeneous Lorentz group. In the quantum mechanical interpretation, however, all the states of the corresponding particles are invariant under translations and, in particular, independent of time. It in very unlikely that these representations have immediate physical significance. Is addition, the third paper contains a determination of those representations for which the momentum vectors are space like. These are not considered in the present article as they also are unlikely to have a simple physical interpretation.

[6] Lubánski, J. K., *Physica*, **IX**, 310–324 (1942).

[7] Dirac, P. A. M., *Proc. Roy. Soc.*, **A, 155,** 447–459 (1936).

[8] The literature on relativistic wave equations is very extensive. Besides the papers quoted in reference 11, we only mention the book by de Broglie, L., *Théorie générale des particules à spin* (Paris, 1943), and the following articles which give a systematic account of the subject: Pauli, W., *Rev. Mod. Phys.*, **13,** 203–232 (1941); Bhabha, H. J., *Rev. Mod. Phys.*, **17,** 203–209 (1945); Kramers, H. A., Belinfante, F. J., and Lubánski, J. K., *Physica*, **VIII,** 597–627 (1941). In this paper, the sum of (14) over all ν was postulated; (14a) then has to be added as an independent equation (except for $N = 1$). Reference 11 uses these equations in the form given by Kramers, Belinfante and Lubánski.

[9] One may derive this result in a more elegant way, without specializing the coördinate system. For the sake of brevity, we omit this derivation.

[10] de Wet, J. S., *Phys. Rev.*, **58,** 236–242 (1940), in particular, p. 242.

[11] Wigner, E. P., *Z. Physik*, (1947).

*STEREOSCOPIC ACUITY FOR VARIOUS LEVELS OF ILLUMINATION**

BY C. G. MUELLER AND V. V. LLOYD

PSYCHOLOGICAL LABORATORY, COLUMBIA UNIVERSITY

Communicated by C. H. Graham, March 5, 1948

Several experiments have demonstrated that the threshold for stereoscopic vision is influenced by certain important variables (see the review by Graham[1]), but little attention has been paid to the systematic exploration of parameters (e.g., intensity and wave-length) which are known to be important for other visual functions.[2] The present report gives data on one of those variables, intensity of "white" light, as it influences the threshold for stereoscopic vision.

Apparatus.—Two 300-watt Mazda bulbs are used as light sources, one for each eye. The light sources are fastened to a movable wooden stand which may be placed in either of two positions, thus allowing for a small range of intensity variation. Additional adjustment of intensity may be achieved by inserting filters of various densities in a holder adjacent to the light source for each eye.

The two filter holders are attached to a pair of metal funnels, 4 inches in diameter and $3^3/_4$ inches in length; the funnels in turn are fastened to the outer wall of the dark room in which the subject sits.

A piece of opal glass, fastened to the inner wall of the dark room and in front of the funnels, diffuses the light from the two bulbs. A piece of masonite, containing two holes of $3^1/_2$ inch diameter, is mounted in front of the opal glass. These holes expose two photographic plates which are fitted into slots in the masonite and on which the reticles of the two visual fields are photographed. Both test fields contain three vertical reticle marks, each with a width of 20 minutes and a height of two degrees of visual angle. The reticle marks are equidistant from one another at a

separation of 4.08 degrees. The central mark is centered in the visual field.

On the front side of the piece of masonite, a vertical target (long vertical line) is permanently fastened over the left hold and to the left of the central reticle mark. The vertical target on the right side, on the contrary, is movable and may be adjusted to the left or the right of its central position by means of a micrometer. Both of the targets have a width of 20 minutes and a length of 19 degrees of visual angle. The micrometer, which the subject uses for adjusting the position of the target, is connected with a metal pointer through the side wall of the dark room; the rotation of the pointer may be read on a degree scale by the experimenter. Each degree of rotation moves the right target through 0.95 second of arc.

A front-silvered mirror is placed five inches in front of each reticle plate to form an angle of 45 degrees with the optical path. A second front-silvered mirror, in both the left and right eye systems, is placed on a movable platform four inches from the first mirror. Each of the latter mirrors also forms an angle of 45 degrees with its optical paths. The two platforms are attached to two eyepieces whose separation may be adjusted to the interpupillary distance of the subject.

All of the above apparatus is enclosed in a large box from which the eyepieces project. Each eyepiece has an aperture of $^3/_8$ inch diameter and holds a convex lens, the focal length of which (14 inches) equals the distance from the lens to the reticle.

The dark room in which the subject is seated is painted a dull black and is completely light tight.

Procedure.—At the start of each experimental session, the subject was allowed to dark adapt for 25 minutes. During the first 15 minutes of this period the subject wore goggles which transmitted the far red region of the spectrum;[3] the last ten minutes were spent in the completely darkened room.

Two subjects were used and each provided three complete sets of data. Both subjects were highly experienced in making stereoscopic settings. Each set of data for each subject consisted of 20 readings taken at each of ten intensity levels ranging from −4.04 log millilamberts to 2.27 log millilamberts. The intensity levels were presented to the subject in order of increasing magnitude and two minutes of light adaptation were given at each level. A rest period, 10 to 20 minutes in length, was given between the fifth and sixth intensity levels in each session.

For each reading the subject started from a randomly selected micrometer position and adjusted the movable target, turning the micrometer until the fused target line appeared in the same plane as the three reticle lines. Both subjects used a "bracketing" procedure in making settings.

The position of the movable target at each adjustment was recorded by the experimenter from the scale on the outside wall of the dark room. Readings at each intensity level were made at the rate of about 4 per minute.

If s_l be the separation between the reference reticle line and target for the left eye and s_r the corresponding separation in the right eye, then the difference, d (i.e., $s_l - s_r$), divided by R, the distance from the lens to the plane of the target and reticles, gives the angular disparity, η, in radians.[1] For threshold,

$$\eta_t = 206{,}265 \frac{d_t}{R} \qquad (1)$$

in seconds of arc, where d_t and R are measured in the same units. In the present experiment, d_t is defined as the average deviation of the "equality" settings.

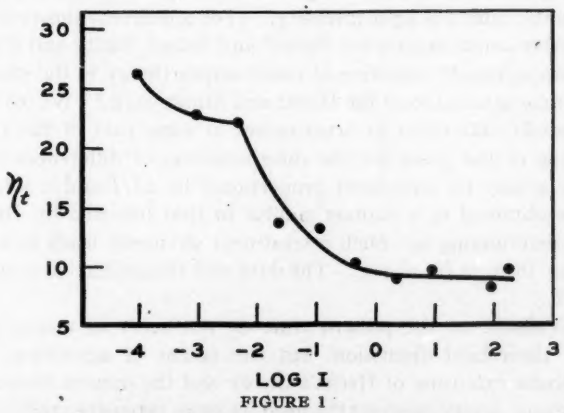

FIGURE 1

The threshold for stereoscopic vision in seconds as a function of field brightness in millilamberts.

Results.—The results[4] of the experiment are presented in figure 1. The data show that η_t decreases (stereoscopic acuity increases) as light intensity increases, until at high intensities, the curve approaches a final limiting value. The mean of the average deviations (representing 120 observations at each intensity) undergoes a threefold change as intensity is changed by a factor of one million. At low intensities the results indicate the usual discontinuity of rod and cone functions.[2] The line drawn through the data is a theoretical curve based on an extension of the Hecht and Mintz[5] formulation of the photochemical basis of visual acuity.

Our data provide some slight indication that the mean setting also varies with the intensity level. The mean settings show considerable variability

but the trend is in the direction of "nearer" mean settings with an increase in intensity.

Discussion.—A theory of stereoscopic vision in terms of its relation to other visual functions is relatively undeveloped.

In experiments on monocular visual acuity for thin lines, Hecht and Mintz[5] found that their results could be accounted for by assuming that the threshold visual angle subtended by the thin lines is proportional to a threshold brightness difference. Their development involved setting the threshold visual angle, α, proportional to $\Delta I/I$ and substituting α in the theoretical solution for $\Delta I/I$ as a function of intensity. The procedure leads to the expression:

$$\alpha = b\left[1 + \frac{1}{(KI)^{1/2}}\right]^2, \qquad (2)$$

where α is the threshold visual angle for resolution of thin lines; b and K are constants; and I is light intensity. (For a different theoretical treatment of other acuity figures see Shlaer[6] and Shlaer, Smith and Chase.[7])

A first approximate extension of visual acuity theory to the stereoscopic situation may proceed from the Hecht and Mintz work.[5] We may assume that threshold differences in brightnesses in some part of the two fields provide one of the bases for the discrimination of differences in depth. Therefore η_t may be considered proportional to $\Delta I/I$ and a solution for η_t may be obtained in a manner similar to that followed by Hecht and Mintz in determining α. Such a treatment obviously leads to a formula of the same form as (2) above. The data and theoretical curve are shown in figure 1.

Data available at the present time do not seem to warrant a more elaborate theoretical discussion, but the extent of agreement between this particular extension of Hecht's theory and the present measurements of stereoscopic acuity suggests the need for more extensive study of parameters of stereoscopic acuity.

Summary.—(1) Measurements of stereoscopic acuity were made at ten levels of intensity ranging from −4.04 to 2.27 log millilamberts. (2) The average deviations of the "equality" settings is taken to be the measure of the minimum resolvable difference angle, η_t, for stereoscopic acuity. η_t is large at low intensities and decreases at high intensities. (3) The typical "rod-cone" discontinuity characteristic of other visual functions is present in the data for stereoscopic acuity. Discrimination of depth seems to be possible at intensities below cone threshold. (4) The results are discussed in terms of theories of brightness discrimination and visual acuity.

* This account was prepared under Contract Number N6onr-271, Task Order IX, between Columbia University and the Office of Naval Research, U. S. Navy. Project NR 142–404.

[1] Graham, C. H., "Visual Perception." In *Handbook of Experimental Psychology* (edited by Stevens). In preparation.

[2] Hecht, S., *Physiol. Rev.*, **17**, 239 (1937).

[3] Miles, W. R., *Fed. Proc.*, **2**, 109 (1943).

[4] The data for the figure are represented by the following pairs of numbers. The first number of each pair is the logarithm of the intensity (in millilamberts); the second the mean of the 6 average deviations (in seconds of arc). 2.27, 9.78; 1.96, 8.34; 0.96, 9.66; 0.36, 9.06; −0.34, 10.43; 0.94, 13.37; −1.64, 13.67; −2.34, 22.17; −3.04, 22.82; −4.04, 26.08.

[5] Hecht, S., and Mintz, E. U., *J. Gen. Physiol.*, **22**, 593 (1939).

[6] Shlaer, S., *J. Gen. Physiol.*, **20**, 165 (1937).

[7] Shlaer, S., Smith, E. L., and Chase, A. M., *J. Gen. Physiol.*, **25**, 553 (1942).